



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12N 15/54, 15/82, 9/10, 5/00, C12P 17/06		A2	(11) International Publication Number: WO 00/63391 (43) International Publication Date: 26 October 2000 (26.10.00)
(21) International Application Number: PCT/US00/10368 (22) International Filing Date: 14 April 2000 (14.04.00) (30) Priority Data: 60/129,899 15 April 1999 (15.04.99) US 60/146,461 30 July 1999 (30.07.99) US (71) Applicant (for all designated States except US): CALGENE LLC [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): SAVIDGE, Beth [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). LASSNER, Michael, W. [US/US]; 1920 Fifth Street, Davis, CA 95616 (US). WEISS, James, D. [US/US]; 800 N. Lindbergh Blvd., St. Louis, MO 63167 (US). POST-BEITTENMILLER, Dusty [US/US]; 800 N. Lindbergh Blvd., St. Louis, MO 63167 (US). (74) Agents: SCHWEDLER, Carl, J. et al.; Calgene LLC, 1920 Fifth Street, Davis, CA 95616 (US).		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>	
(54) Title: NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS			
(57) Abstract <p>Nucleic acid sequences and methods are provided for producing plants and seeds having altered tocopherol content and compositions. The methods find particular use in increasing the tocopherol levels in plants, and in providing desirable tocopherol compositions in a host plant cell.</p>			
Della Penna, et al. S/N 09/560,761			R A

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

NUCLEIC ACID SEQUENCES TO PROTEINS INVOLVED IN TOCOPHEROL SYNTHESIS

5

INTRODUCTION

This application claims the benefit of the filing date of the provisional Application U.S. Serial Number 60/129,899, filed April 15, 1999, and the provisional Application, U.S.
10 Serial Number 60/146,461, filed July 30, 1999.

TECHNICAL FIELD

The present invention is directed to nucleic acid and amino acid sequences and constructs, and methods related thereto.

15

BACKGROUND

Isoprenoids are ubiquitous compounds found in all living organisms. Plants synthesize a diverse array of greater than 22,000 isoprenoids (Connolly and Hill (1992) *Dictionary of Terpenoids*, Chapman and Hall, New York, NY). In plants, isoprenoids play essential roles in
20 particular cell functions such as production of sterols, contributing to eukaryotic membrane architecture, acyclic polyprenoids found in the side chain of ubiquinone and plastoquinone, growth regulators like abscisic acid, gibberellins, brassinosteroids or the photosynthetic pigments chlorophylls and carotenoids. Although the physiological role of other plant isoprenoids is less evident, like that of the vast array of secondary metabolites, some are
25 known to play key roles mediating the adaptative responses to different environmental challenges. In spite of the remarkable diversity of structure and function, all isoprenoids originate from a single metabolic precursor, isopentenyl diphosphate (IPP) (Wright, (1961) *Annu. Rev. Biochem.* 20:525-548; and Spurgeon and Porter, (1981) in Biosynthesis of Isoprenoid Compounds, Porter and Spurgeon eds (John Wiley, New York) Vol. 1, pp1-46).

30

A number of unique and interconnected biochemical pathways derived from the isoprenoid pathway leading to secondary metabolites, including tocopherols, exist in chloroplasts of higher plants. Tocopherols not only perform vital functions in plants, but are

also important from mammalian nutritional perspectives. In plastids, tocopherols account for up to 40% of the total quinone pool.

5 Tocopherols and tocotrienols (unsaturated tocopherol derivatives) are well known antioxidants, and play an important role in protecting cells from free radical damage, and in the prevention of many diseases, including cardiac disease, cancer, cataracts, retinopathy, Alzheimer's disease, and neurodegeneration, and have been shown to have beneficial effects on symptoms of arthritis, and in anti-aging. Vitamin E is used in chicken feed for improving the shelf life, appearance, flavor, and oxidative stability of meat, and to transfer tocopherols from feed to eggs. Vitamin E has been shown to be essential for normal reproduction, improves overall performance, and enhances immunocompetence in livestock animals. Vitamin E supplement in animal feed also imparts oxidative stability to milk products.

10 The demand for natural tocopherols as supplements has been steadily growing at a rate of 10-20% for the past three years. At present, the demand exceeds the supply for natural tocopherols, which are known to be more biopotent than racemic mixtures of synthetically produced tocopherols. Naturally occurring tocopherols are all *d*-stereoisomers, whereas synthetic α -tocopherol is a mixture of eight *d,l*- α -tocopherol isomers, only one of which (12.5%) is identical to the natural *d*- α -tocopherol. Natural *d*- α -tocopherol has the highest vitamin E activity (1.49 IU/mg) when compared to other natural tocopherols or tocotrienols. The synthetic α -tocopherol has a vitamin E activity of 1.1 IU/mg. In 1995, the worldwide market for raw refined tocopherols was \$1020 million; synthetic materials comprised 85-88% of the market, the remaining 12-15% being natural materials. The best sources of natural tocopherols and tocotrienols are vegetable oils and grain products. Currently, most of the natural Vitamin E is produced from γ -tocopherol derived from soy oil processing, which is subsequently converted to α -tocopherol by chemical modification (α -tocopherol exhibits the greatest biological activity).

20 Methods of enhancing the levels of tocopherols and tocotrienols in plants, especially levels of the more desirable compounds that can be used directly, without chemical modification, would be useful to the art as such molecules exhibit better functionality and bioavailability.

30 In addition, methods for the increased production of other isoprenoid derived compounds in a host plant cell is desirable. Furthermore, methods for the production of particular isoprenoid compounds in a host plant cell is also needed.

SUMMARY OF THE INVENTION

5 The present invention is directed to prenyltransferase (PT), and in particular to PT polynucleotides and polypeptides. The polynucleotides and polypeptides of the present invention include those derived from prokaryotic and eukaryotic sources.

 Thus, one aspect of the present invention relates to isolated polynucleotide sequences encoding prenyltransferase proteins. In particular, isolated nucleic acid sequences encoding
10 PT proteins from bacterial and plant sources are provided.

 Another aspect of the present invention relates to oligonucleotides which include partial or complete PT encoding sequences.

 It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of
15 prenyltransferase. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells.

 In another aspect of the present invention, methods are provided for production of prenyltransferase in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and
20 translation of prenyltransferase. The recombinant cells which contain prenyltransferase are also part of the present invention.

 In a further aspect, the present invention relates to methods of using polynucleotide and polypeptide sequences to modify the tocopherol content of host cells, particularly in host plant cells. Plant cells having such a modified tocopherol content are also contemplated
25 herein.

 The modified plants, seeds and oils obtained by the expression of the prenyltransferases are also considered part of the invention.

BRIEF DESCRIPTION OF THE DRAWINGS

30

 Figure 1 provides an amino acid sequence alignment between ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 are performed using ClustalW.

Figure 2 provides a schematic picture of the expression construct pCGN10800.

Figure 3 provides a schematic picture of the expression construct pCGN10801.

Figure 4 provides a schematic picture of the expression construct pCGN10803.

Figure 5 provides a schematic picture of the expression construct pCGN10806.

5 Figure 6 provides a schematic picture of the expression construct pCGN10807.

Figure 7 provides a schematic picture of the expression construct pCGN10808.

Figure 8 provides a schematic picture of the expression construct pCGN10809.

Figure 9 provides a schematic picture of the expression construct pCGN10810.

Figure 10 provides a schematic picture of the expression construct pCGN10811.

10 Figure 11 provides a schematic picture of the expression construct pCGN10812.

Figure 12 provides a schematic picture of the expression construct pCGN10813.

Figure 13 provides a schematic picture of the expression construct pCGN10814.

Figure 14 provides a schematic picture of the expression construct pCGN10815.

Figure 15 provides a schematic picture of the expression construct pCGN10816.

15 Figure 16 provides a schematic picture of the expression construct pCGN10817.

Figure 17 provides a schematic picture of the expression construct pCGN10819.

Figure 18 provides a schematic picture of the expression construct pCGN10824.

Figure 19 provides a schematic picture of the expression construct pCGN10825.

Figure 20 provides a schematic picture of the expression construct pCGN10826.

20 Figure 21 provides an amino acid sequence alignment using ClustalW between the *Synechocystis* sequence knockouts.

Figure 22 provides an amino acid sequence of the ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 protein sequences from *Arabidopsis* and the slr1736, slr0926, slr11899, slr0056, and the slr1518 amino acid sequences from *Synechocystis*.

25 Figure 23 provides the results of the enzymatic assay from preparations of wild type *Synechocystis* strain 6803, and *Synechocystis* slr1736 knockout.

Figure 24 provides bar graphs of HPLC data obtained from seed extracts of transgenic *Arabidopsis* containing pCGN10822, which provides of the expression of the ATPT2 sequence, in the sense orientation, from the napin promoter. Provided are graphs for alpha, gamma, and delta tocopherols, as well as total tocopherol for 22 transformed lines, as well as a nontransformed (wildtype) control.

30

Figure 25 provides a bar graph of HPLC analysis of seed extracts from *Arabidopsis* plants transformed with pCGN10803 (35S-ATPT2, in the antisense orientation), pCGN10802

(line 1625, napin ATPT2 in the sense orientation), pCGN10809 (line 1627, 35S-ATPT3 in the sense orientation), a nontransformed (wt) control, and a empty vector transformed control.

5

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides, *inter alia*, compositions and methods for altering (for example, increasing and decreasing) the tocopherol levels and/or modulating their ratios in
10 host cells. In particular, the present invention provides polynucleotides, polypeptides, and methods of use thereof for the modulation of tocopherol content in host plant cells.

The present invention provides polynucleotide and polypeptide sequences involved in the prenylation of straight chain and aromatic compounds. Straight chain prenyl transferases as used herein comprises sequences which encode proteins involved in the prenylation of
15 straight chain compounds, including, but not limited to, geranyl geranyl pyrophosphate and farnesyl pyrophosphate. Aromatic prenyl transferases, as used herein, comprises sequences which encode proteins involved in the prenylation of aromatic compounds, including, but not limited to, menaquinone, ubiquinone, chlorophyll, and homogentisic acid. The prenyl transferase of the present invention preferably prenylates homogentisic acid.

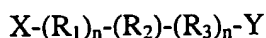
20 The biosynthesis of α -tocopherol in higher plants involves condensation of homogentisic acid and phytylpyrophosphate to form 2-methyl-6-phytylbenzoquinol that can, by cyclization and subsequent methylations (Fiedler et al., 1982, *Planta*, 155: 511-515, Soll et al., 1980, *Arch. Biochem. Biophys.* 204: 544-550, Marshall et al., 1985 *Phytochem.*, 24: 1705-1711, all of which are herein incorporated by reference in their entirety), form various
25 tocopherols. The *Arabidopsis pds2* mutant identified and characterized by Norris *et al.* (1995), is deficient in tocopherol and plastoquinone-9 accumulation. Further genetic and biochemical analysis suggests that the protein encoded by *PDS2* may be responsible for the prenylation of homogentisic acid. This may be a rate limiting step in tocopherol biosynthesis, and this gene has yet to be isolated. Thus, it is an aspect of the present invention to provide
30 polynucleotides and polypeptides involved in the prenylation of homogentisic acid.

Isolated Polynucleotides, Proteins, and Polypeptides

A first aspect of the present invention relates to isolated prenyltransferase polynucleotides. The polynucleotide sequences of the present invention include isolated polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing and to other polynucleotide sequences closely related to such sequences and variants thereof.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

The invention also includes polynucleotides of the formula:



wherein, at the 5' end, X is hydrogen, and at the 3' end, Y is hydrogen or a metal, R_1 and R_3 are any nucleic acid residue, n is an integer between 1 and 3000, preferably between 1 and 1000 and R_2 is a nucleic acid sequence of the invention, particularly a nucleic acid sequence selected from the group set forth in the Sequence Listing and preferably those of SEQ ID NOs: 1, 3, 5, 7, 8, 10, 11, 13-16, 18, 23, 29, 36, and 38. In the formula, R_2 is oriented so that its 5' end residue is at the left, bound to R_1 , and its 3' end residue is at the right, bound to R_3 . Any stretch of nucleic acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

The invention also relates to variants of the polynucleotides described herein that encode for variants of the polypeptides of the invention. Variants that are fragments of the polynucleotides of the invention can be used to synthesize full-length polynucleotides of the invention. Preferred embodiments are polynucleotides encoding polypeptide variants wherein

5 to 10, 1 to 5, 1 to 3, 2, 1 or no amino acid residues of a polypeptide sequence of the invention are substituted, added or deleted, in any combination. Particularly preferred are substitutions, additions, and deletions that are silent such that they do not alter the properties or activities of the polynucleotide or polypeptide.

5 Further preferred embodiments of the invention that are at least 50%, 60%, or 70% identical over their entire length to a polynucleotide encoding a polypeptide of the invention, and polynucleotides that are complementary to such polynucleotides. More preferable are polynucleotides that comprise a region that is at least 80% identical over its entire length to a polynucleotide encoding a polypeptide of the invention and polynucleotides that are
10 complementary thereto. In this regard, polynucleotides at least 90% identical over their entire length are particularly preferred, those at least 95% identical are especially preferred. Further, those with at least 97% identity are highly preferred and those with at least 98% and 99% identity are particularly highly preferred, with those at least 99% being the most highly preferred.

15 Preferred embodiments are polynucleotides that encode polypeptides that retain substantially the same biological function or activity as the mature polypeptides encoded by the polynucleotides set forth in the Sequence Listing.

The invention further relates to polynucleotides that hybridize to the above-described sequences. In particular, the invention relates to polynucleotides that hybridize under
20 stringent conditions to the above-described polynucleotides. As used herein, the terms "stringent conditions" and "stringent hybridization conditions" mean that hybridization will generally occur if there is at least 95% and preferably at least 97% identity between the sequences. An example of stringent hybridization conditions is overnight incubation at 42°C in a solution comprising 50% formamide, 5x SSC (150 mM NaCl, 15 mM trisodium citrate),
25 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 micrograms/milliliter denatured, sheared salmon sperm DNA, followed by washing the hybridization support in 0.1x SSC at approximately 65°C. Other hybridization and wash conditions are well known and are exemplified in Sambrook, *et al.*, Molecular Cloning: A Laboratory Manual, Second Edition, Cold Spring Harbor, NY (1989), particularly Chapter 11.

30 The invention also provides a polynucleotide consisting essentially of a polynucleotide sequence obtainable by screening an appropriate library containing the complete gene for a polynucleotide sequence set forth in the Sequence Listing under stringent hybridization conditions with a probe having the sequence of said polynucleotide sequence or

a fragment thereof; and isolating said polynucleotide sequence. Fragments useful for obtaining such a polynucleotide include, for example, probes and primers as described herein.

As discussed herein regarding polynucleotide assays of the invention, for example, polynucleotides of the invention can be used as a hybridization probe for RNA, cDNA, or genomic DNA to isolate full length cDNAs or genomic clones encoding a polypeptide and to isolate cDNA or genomic clones of other genes that have a high sequence similarity to a polynucleotide set forth in the Sequence Listing. Such probes will generally comprise at least 15 bases. Preferably such probes will have at least 30 bases and can have at least 50 bases. Particularly preferred probes will have between 30 bases and 50 bases, inclusive.

The coding region of each gene that comprises or is comprised by a polynucleotide sequence set forth in the Sequence Listing may be isolated by screening using a DNA sequence provided in the Sequence Listing to synthesize an oligonucleotide probe. A labeled oligonucleotide having a sequence complementary to that of a gene of the invention is then used to screen a library of cDNA, genomic DNA or mRNA to identify members of the library which hybridize to the probe. For example, synthetic oligonucleotides are prepared which correspond to the prenyltransferase EST sequences. The oligonucleotides are used as primers in polymerase chain reaction (PCR) techniques to obtain 5' and 3' terminal sequence of prenyl transferase genes. Alternatively, where oligonucleotides of low degeneracy can be prepared from particular prenyltransferase peptides, such probes may be used directly to screen gene libraries for prenyltransferase gene sequences. In particular, screening of cDNA libraries in phage vectors is useful in such methods due to lower levels of background hybridization.

Typically, a prenyltransferase sequence obtainable from the use of nucleic acid probes will show 60-70% sequence identity between the target prenyltransferase sequence and the encoding sequence used as a probe. However, lengthy sequences with as little as 50-60% sequence identity may also be obtained. The nucleic acid probes may be a lengthy fragment of the nucleic acid sequence, or may also be a shorter, oligonucleotide probe. When longer nucleic acid fragments are employed as probes (greater than about 100 bp), one may screen at lower stringencies in order to obtain sequences from the target sample which have 20-50% deviation (i.e., 50-80% sequence homology) from the sequences used as probe.

Oligonucleotide probes can be considerably shorter than the entire nucleic acid sequence encoding an prenyltransferase enzyme, but should be at least about 10, preferably at least about 15, and more preferably at least about 20 nucleotides. A higher degree of sequence

identity is desired when shorter regions are used as opposed to longer regions. It may thus be desirable to identify regions of highly conserved amino acid sequence to design oligonucleotide probes for detecting and recovering other related prenyltransferase genes. Shorter probes are often particularly useful for polymerase chain reactions (PCR), especially when highly conserved sequences can be identified. (See, Gould, *et al.*, *PNAS USA* (1989) 86:1934-1938.).

Another aspect of the present invention relates to prenyltransferase polypeptides. Such polypeptides include isolated polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit prenyltransferase activity and also those polypeptides which have at least 50%, 60% or 70% identity, preferably at least 80% identity, more preferably at least 90% identity, and most preferably at least 95% identity to a polypeptide sequence selected from the group of sequences set forth in the Sequence Listing, and also include portions of such polypeptides, wherein such portion of the polypeptide preferably includes at least 30 amino acids and more preferably includes at least 50 amino acids.

"Identity", as is well understood in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as determined by the match between strings of such sequences. "Identity" can be readily calculated by known methods including, but not limited to, those described in *Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part I*, Griffin, A.M. and Griffin, H.G., eds., Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press (1987); *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., Stockton Press, New York (1991); and Carillo, H., and Lipman, D., *SIAM J Applied Math*, 48:1073 (1988). Methods to determine identity are designed to give the largest match between the sequences tested. Moreover, methods to determine identity are codified in publicly available programs. Computer programs which can be used to determine identity between two sequences include, but are not limited to, GCG (Devereux, J., et al., *Nucleic Acids Research* 12(1):387 (1984); suite of five BLAST programs, three designed for nucleotide sequences queries (BLASTN, BLASTX, and TBLASTX) and two designed for protein sequence queries (BLASTP and TBLASTN)

(Coulson, *Trends in Biotechnology*, 12: 76-80 (1994); Birren, *et al.*, *Genome Analysis*, 1: 543-559 (1997)). The BLAST X program is publicly available from NCBI and other sources (*BLAST Manual*, Altschul, S., *et al.*, NCBI NLM NIH, Bethesda, MD 20894; Altschul, S., *et al.*, *J. Mol. Biol.*, 215:403-410 (1990)). The well known Smith Waterman algorithm can also
 5 be used to determine identity.

Parameters for polypeptide sequence comparison typically include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

Comparison matrix: BLOSSUM62 from Hentikoff and Hentikoff, *Proc. Natl. Acad. Sci USA* 89:10915-10919 (1992)

10 Gap Penalty: 12

Gap Length Penalty: 4

A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters along with no penalty for end gap are the default parameters for peptide comparisons.

15 Parameters for polynucleotide sequence comparison include the following:

Algorithm: Needleman and Wunsch, *J. Mol. Biol.* 48:443-453 (1970)

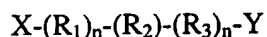
Comparison matrix: matches = +10; mismatches = 0

Gap Penalty: 50

Gap Length Penalty: 3

20 A program which can be used with these parameters is publicly available as the "gap" program from Genetics Computer Group, Madison Wisconsin. The above parameters are the default parameters for nucleic acid comparisons.

The invention also includes polypeptides of the formula:



25 wherein, at the amino terminus, X is hydrogen, and at the carboxyl terminus, Y is hydrogen or a metal, R_1 and R_3 are any amino acid residue, n is an integer between 1 and 1000, and R_2 is an amino acid sequence of the invention, particularly an amino acid sequence selected from the group set forth in the Sequence Listing and preferably those encoded by the sequences provided in SEQ ID NOs: 2, 4, 6, 9, 12, 17, 19-22, 24-28, 30, 32-35, 37, and 39. In the
 30 formula, R_2 is oriented so that its amino terminal residue is at the left, bound to R_1 , and its carboxy terminal residue is at the right, bound to R_3 . Any stretch of amino acid residues denoted by either R group, where R is greater than 1, may be either a heteropolymer or a homopolymer, preferably a heteropolymer.

Polypeptides of the present invention include isolated polypeptides encoded by a polynucleotide comprising a sequence selected from the group of a sequence contained in the Sequence Listing set forth herein .

5 The polypeptides of the present invention can be mature protein or can be part of a fusion protein.

Fragments and variants of the polypeptides are also considered to be a part of the invention. A fragment is a variant polypeptide which has an amino acid sequence that is entirely the same as part but not all of the amino acid sequence of the previously described polypeptides. The fragments can be "free-standing" or comprised within a larger polypeptide
10 of which the fragment forms a part or a region, most preferably as a single continuous region. Preferred fragments are biologically active fragments which are those fragments that mediate activities of the polypeptides of the invention, including those with similar activity or improved activity or with a decreased activity. Also included are those fragments that antigenic or immunogenic in an animal, particularly a human.

15 Variants of the polypeptide also include polypeptides that vary from the sequences set forth in the Sequence Listing by conservative amino acid substitutions, substitution of a residue by another with like characteristics. In general, such substitutions are among Ala, Val, Leu and Ile; between Ser and Thr; between Asp and Glu; between Asn and Gln; between Lys and Arg; or between Phe and Tyr. Particularly preferred are variants in which 5 to 10; 1
20 to 5; 1 to 3 or one amino acid(s) are substituted, deleted, or added, in any combination.

Variants that are fragments of the polypeptides of the invention can be used to produce the corresponding full length polypeptide by peptide synthesis. Therefore, these variants can be used as intermediates for producing the full-length polypeptides of the invention.

25 The polynucleotides and polypeptides of the invention can be used, for example, in the transformation of host cells, such as plant host cells, as further discussed herein.

The invention also provides polynucleotides that encode a polypeptide that is a mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids within the mature polypeptide (for example, when the mature form of the protein has more than one
30 polypeptide chain). Such sequences can, for example, play a role in the processing of a protein from a precursor to a mature form, allow protein transport, shorten or lengthen protein half-life, or facilitate manipulation of the protein in assays or production. It is contemplated

that cellular enzymes can be used to remove any additional amino acids from the mature protein.

A precursor protein, having the mature form of the polypeptide fused to one or more prosequences may be an inactive form of the polypeptide. The inactive precursors generally are activated when the prosequences are removed. Some or all of the prosequences may be removed prior to activation. Such precursor protein are generally called proproteins.

Plant Constructs and Methods of Use

Of particular interest is the use of the nucleotide sequences in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the prenyltransferase sequences of the present invention in a host plant cell. The expression constructs generally comprise a promoter functional in a host plant cell operably linked to a nucleic acid sequence encoding a prenyltransferase of the present invention and a transcriptional termination region functional in a host plant cell.

A first nucleic acid sequence is "operably linked" or "operably associated" with a second nucleic acid sequence when the sequences are so arranged that the first nucleic acid sequence affects the function of the second nucleic-acid sequence. Preferably, the two sequences are part of a single contiguous nucleic acid molecule and more preferably are adjacent. For example, a promoter is operably linked to a gene if the promoter regulates or mediates transcription of the gene in a cell.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of plant functional promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the prenyltransferase gene in specific tissues of the plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res.* 1:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearyl-ACP desaturase, soybean α' subunit of β -conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins conferring prenyltransferase to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the gene of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the gene of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep.* 9:104-126; Clark *et al.* (1989) *J. Biol. Chem.* 264:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol.* 84:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun.* 196:1414-1421; and, Shah *et al.* (1986) *Science* 233:478-481.

Depending upon the intended use, the constructs may contain the nucleic acid sequence which encodes the entire prenyltransferase protein, or a portion thereof. For example, where antisense inhibition of a given prenyltransferase protein is desired, the entire prenyltransferase sequence is not required. Furthermore, where prenyltransferase sequences used in constructs are intended for use as probes, it may be advantageous to prepare constructs containing only a particular portion of a prenyltransferase encoding sequence, for example a sequence which is discovered to encode a highly conserved prenyltransferase region.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to, antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726), co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the prenyltransferase or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the prenyltransferase sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

The prenyltransferase constructs of the present invention can be used in transformation methods with additional constructs providing for the expression of other nucleic acid sequences encoding proteins involved in the production of tocopherols, or tocopherol precursors such as homogentisic acid and/or phytylpyrophosphate. Nucleic acid sequences encoding proteins involved in the production of homogentisic acid are known in the art, and include but are not limited to, 4-hydroxyphenylpyruvate dioxygenase (HPPD, EC 1.13.11.27) described for example, by Garcia, *et al.* ((1999) *Plant Physiol.* 119(4):1507-1516), mono or bifunctional tyrA (described for example by Xia, *et al.* (1992) *J. Gen Microbiol.* 138:1309-1316, and Hudson, *et al.* (1984) *J. Mol. Biol.* 180:1023-1051), Oxygenase, 4-hydroxyphenylpyruvate di- (9CI), 4-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate dioxygenase; p-Hydroxyphenylpyruvate hydroxylase; p-Hydroxyphenylpyruvate oxidase; p-Hydroxyphenylpyruvic acid hydroxylase; p-

Hydroxyphenylpyruvic hydroxylase; p-Hydroxyphenylpyruvic oxidase), 4-hydroxyphenylacetate, NAD(P)H: oxygen oxidoreductase (1-hydroxylating); 4-hydroxyphenylacetate 1-monooxygenase, and the like. In addition, constructs for the expression of nucleic acid sequences encoding proteins involved in the production of
5 phytylpyrophosphate can also be employed with the prenyltransferase constructs of the present invention. Nucleic acid sequences encoding proteins involved in the production of phytylpyrophosphate are known in the art, and include, but are not limited to geranylgeranylpyrophosphate synthase (GGPPS), geranylgeranylpyrophosphate reductase (GGH), 1-deoxyxylulose-5-phosphate synthase, 1-deoxy-D-xylulose-5-phosphate
10 reductoisomerase, 4-diphosphocytidyl-2-C-methylerythritol synthase, isopentyl pyrophosphate isomerase.

The prenyltransferase sequences of the present invention find use in the preparation of transformation constructs having a second expression cassette for the expression of additional sequences involved in tocopherol biosynthesis. Additional tocopherol biosynthesis
15 sequences of interest in the present invention include, but are not limited to gamma-tocopherol methyltransferase (Shintani, *et al.* (1998) *Science* 282(5396):2098-2100), tocopherol cyclase, and tocopherol methyltransferase.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed,
20 transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a prenyltransferase nucleic acid sequence.

Plant expression or transcription constructs having a prenyltransferase as the DNA
25 sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible and industrial uses. Particularly preferred plants for use in the methods of the present invention include, but are not limited to: *Acacia*, alfalfa, aneth, apple, apricot, artichoke, arugula, asparagus, avocado, banana, barley, beans, beet, blackberry, blueberry,
30 broccoli, brussels sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, celery, cherry, chicory, cilantro, citrus, clementines, coffee, corn, cotton, cucumber, Douglas fir, eggplant, endive, escarole, eucalyptus, fennel, figs, garlic, gourd, grape, grapefruit, honey dew, jicama, kiwifruit, lettuce, leeks, lemon, lime, Loblolly pine, mango, melon, mushroom,

nectarine, nut, oat, oil palm, oil seed rape, okra, onion, orange, an ornamental plant, papaya, parsley, pea, peach, peanut, pear, pepper, persimmon, pine, pineapple, plantain, plum, pomegranate, poplar, potato, pumpkin, quince, radiata pine, radicchio, radish, raspberry, rice, rye, sorghum, Southern pine, soybean, spinach, squash, strawberry, sugarbeet, sugarcane, sunflower, sweet potato, sweetgum, tangerine, tea, tobacco, tomato, triticale, turf, turnip, a vine, watermelon, wheat, yams, and zucchini.

Most especially preferred are temperate oilseed crops. Temperate oilseed crops of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of prenyltransferase constructs in plants to produce plants or plant parts, including, but not limited to leaves, stems, roots, reproductive, and seed, with a modified content of tocopherols in plant parts having transformed plant cells.

For immunological screening, antibodies to the protein can be prepared by injecting rabbits or mice with the purified protein or portion thereof, such methods of preparing antibodies being well known to those in the art. Either monoclonal or polyclonal antibodies can be produced, although typically polyclonal antibodies are more useful for gene isolation. Western analysis may be conducted to determine that a related protein is present in a crude extract of the desired plant species, as determined by cross-reaction with the antibodies to the encoded proteins. When cross-reactivity is observed, genes encoding the related proteins are isolated by screening expression libraries representing the desired plant species. Expression libraries can be constructed in a variety of commercially available vectors, including lambda gt11, as described in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual*, Second Edition (1989) Cold Spring Harbor Laboratory, Cold Spring Harbor, New York).

To confirm the activity and specificity of the proteins encoded by the identified nucleic acid sequences as prenyltransferase enzymes, *in vitro* assays are performed in insect cell cultures using baculovirus expression systems. Such baculovirus expression systems are known in the art and are described by Lee, *et al.* U.S. Patent Number 5,348,886, the entirety of which is herein incorporated by reference.

In addition, other expression constructs may be prepared to assay for protein activity utilizing different expression systems. Such expression constructs are transformed into yeast or prokaryotic host and assayed for prenyltransferase activity. Such expression systems are known in the art and are readily available through commercial sources.

5 In addition to the sequences described in the present invention, DNA coding sequences useful in the present invention can be derived from algae, fungi, bacteria, mammalian sources, plants, etc. Homology searches in existing databases using signature sequences corresponding to conserved nucleotide and amino acid sequences of prenyltransferase can be employed to isolate equivalent, related genes from other sources such as plants and
10 microorganisms. Searches in EST databases can also be employed. Furthermore, the use of DNA sequences encoding enzymes functionally enzymatically equivalent to those disclosed herein, wherein such DNA sequences are degenerate equivalents of the nucleic acid sequences disclosed herein in accordance with the degeneracy of the genetic code, is also encompassed by the present invention. Demonstration of the functionality of coding
15 sequences identified by any of these methods can be carried out by complementation of mutants of appropriate organisms, such as *Synechocystis*, *Shewanella*, *yeast*, *Pseudomonas*, *Rhodobacteria*, etc., that lack specific biochemical reactions, or that have been mutated. The sequences of the DNA coding regions can be optimized by gene resynthesis, based on codon usage, for maximum expression in particular hosts.

20 For the alteration of tocopherol production in a host cell, a second expression construct can be used in accordance with the present invention. For example, the prenyltransferase expression construct can be introduced into a host cell in conjunction with a second expression construct having a nucleotide sequence for a protein involved in tocopherol biosynthesis.

25 The method of transformation in obtaining such transgenic plants is not critical to the instant invention, and various methods of plant transformation are currently available. Furthermore, as newer methods become available to transform crops, they may also be directly applied hereunder. For example, many plant species naturally susceptible to *Agrobacterium* infection may be successfully transformed via tripartite or binary vector
30 methods of *Agrobacterium* mediated transformation. In many instances, it will be desirable to have the construct bordered on one or both sides by T-DNA, particularly having the left and right borders, more particularly the right border. This is particularly useful when the construct uses *A. tumefaciens* or *A. rhizogenes* as a mode for transformation, although the T-

DNA borders may find use with other modes of transformation. In addition, techniques of microinjection, DNA particle bombardment, and electroporation have been developed which allow for the transformation of various monocot and dicot plant species.

Normally, included with the DNA construct will be a structural gene having the
5 necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed,
10 where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation)
15 or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s)
20 will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) 77:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector
25 containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride, *et al.* (*Plant Mol. Biol.* (1990) 14:269-276), wherein the pRiHRI (Jouanin, *et al.*, *Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

30 Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed *Agrobacterium* and transformed plant cells. A number of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The

particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the expression construct of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the prenyltransferase expression construct, or alternatively, transformed plants, one expressing the prenyltransferase construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

The nucleic acid sequences of the present invention can be used in constructs to provide for the expression of the sequence in a variety of host cells, both prokaryotic eukaryotic. Host cells of the present invention preferably include monocotyledenous and dicotyledenous plant cells.

In general, the skilled artisan is familiar with the standard resource materials which describe specific conditions and procedures for the construction, manipulation and isolation of macromolecules (e.g., DNA molecules, plasmids, etc.), generation of recombinant organisms and the screening and isolating of clones, (see for example, Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press (1989); Maliga *et al.*, *Methods in Plant Molecular Biology*, Cold Spring Harbor Press (1995), the entirety of which is herein incorporated by reference; Birren *et al.*, *Genome Analysis: Analyzing DNA*, 1, Cold Spring Harbor, New York, the entirety of which is herein incorporated by reference).

Methods for the expression of sequences in insect host cells are known in the art. Baculovirus expression vectors are recombinant insect viruses in which the coding sequence for a chosen foreign gene has been inserted behind a baculovirus promoter in place of the viral gene, e.g., polyhedrin (Smith and Summers, U.S. Pat. No., 4,745,051, the entirety of which is incorporated herein by reference). Baculovirus expression vectors are known in the art, and are described for example in Doerfler, *Curr. Top. Microbiol. Immunol.* 131:51-68 (1968); Luckow and Summers, *Bio/Technology* 6:47-55 (1988a); Miller, *Annual Review of Microbiol.* 42:177-199 (1988); Summers, *Curr. Comm. Molecular Biology*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y. (1988); Summers and Smith, *A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures*, Texas Ag. Exper. Station Bulletin No. 1555 (1988), the entireties of which is herein incorporated by reference)

Methods for the expression of a nucleic acid sequence of interest in a fungal host cell are known in the art. The fungal host cell may, for example, be a yeast cell or a filamentous fungal cell. Methods for the expression of DNA sequences of interest in yeast cells are generally described in "Guide to yeast genetics and molecular biology", Guthrie and Fink, eds. *Methods in enzymology*, Academic Press, Inc. Vol 194 (1991) and *Gene expression technology*, Goeddel ed, *Methods in Enzymology*, Academic Press, Inc., Vol 185 (1991).

Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Suitable promoters for mammalian cells are also known in the art and include, but are not limited to, viral promoters such as that from Simian Virus 40 (SV40) (Fiers *et al.*, *Nature* 273:113 (1978), the entirety of which is herein incorporated by reference), Rous sarcoma virus (RSV), adenovirus (ADV) and bovine papilloma virus (BPV). Mammalian cells may also require terminator sequences and poly-A addition sequences. Enhancer sequences which increase expression may also be included and sequences which promote amplification of the gene may also be desirable (for example methotrexate resistance genes).

Vectors suitable for replication in mammalian cells are well known in the art, and may include viral replicons, or sequences which insure integration of the appropriate sequences encoding epitopes into the host genome. Plasmid vectors that greatly facilitate the construction of recombinant viruses have been described (*see*, for example, Mackett *et al.*, *J Virol.* 49:857 (1984); Chakrabarti *et al.*, *Mol. Cell. Biol.* 5:3403 (1985); Moss, In: *Gene*

Transfer Vectors For Mammalian Cells (Miller and Calos, eds., Cold Spring Harbor Laboratory, N.Y., p. 10, (1987); all of which are herein incorporated by reference in their entirety).

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1: Identification of Prenyltransferase Sequences

PSI-BLAST (Altschul, *et al.* (1997) *Nuc Acid Res* 25:3389-3402) profiles were generated for both the straight chain and aromatic classes of prenyltransferases. To generate the straight chain profile, a prenyl-transferase from *Porphyra purpurea* (Genbank accession 1709766) was used as a query against the NCBI non-redundant protein database. The *E. coli* enzyme involved in the formation of ubiquinone, ubiA (genbank accession 1790473) was used as a starting sequence to generate the aromatic prenyltransferase profile. These profiles were used to search public and proprietary DNA and protein data bases. In *Arabidopsis* seven putative prenyltransferases of the straight-chain class were identified, ATPT1, (SEQ ID NO:9), ATPT7 (SEQ ID NO:10), ATPT8 (SEQ ID NO:11), ATPT9 (SEQ ID NO:13), ATPT10 (SEQ ID NO:14), ATPT11 (SEQ ID NO:15), and ATPT12 (SEQ ID NO:16) and five were identified of the aromatic class, ATPT2 (SEQ ID NO:1), ATPT3 (SEQ ID NO:3), ATPT4 (SEQ ID NO:5), ATPT5 (SEQ ID NO:7), ATPT6 (SEQ ID NO:8). Additional prenyltransferase sequences from other plants related to the aromatic class of prenyltransferases, such as soy (SEQ ID NOs: 19-23, the deduced amino acid sequence of SEQ ID NO:23 is provided in SEQ ID NO:24) and maize (SEQ ID NOs:25-29, and 31) are also identified. The deduced amino acid sequence of ZMPT5 (SEQ ID NO:29) is provided in SEQ ID NO:30.

Searches are performed on a Silicon Graphics Unix computer using additional Bioaccelerator hardware and GenWeb software supplied by Compugen Ltd. This software and hardware enables the use of the Smith-Waterman algorithm in searching DNA and protein databases using profiles as queries. The program used to query protein databases is

profilesearch. This is a search where the query is not a single sequence but a profile based on a multiple alignment of amino acid or nucleic acid sequences. The profile is used to query a sequence data set, i.e., a sequence database. The profile contains all the pertinent information for scoring each position in a sequence, in effect replacing the "scoring matrix" used for the standard query searches. The program used to query nucleotide databases with a protein profile is tprofilesearch. Tprofilesearch searches nucleic acid databases using an amino acid profile query. As the search is running, sequences in the database are translated to amino acid sequences in six reading frames. The output file for tprofilesearch is identical to the output file for profilesearch except for an additional column that indicates the frame in which the best alignment occurred.

The Smith-Waterman algorithm, (Smith and Waterman (1981) *supra*), is used to search for similarities between one sequence from the query and a group of sequences contained in the database. E score values as well as other sequence information, such as conserved peptide sequences are used to identify related sequences.

To obtain the entire coding region corresponding to the *Arabidopsis* prenyltransferase sequences, synthetic oligo-nucleotide primers are designed to amplify the 5' and 3' ends of partial cDNA clones containing prenyltransferase sequences. Primers are designed according to the respective *Arabidopsis* prenyltransferase sequences and used in Rapid Amplification of cDNA Ends (RACE) reactions (Frohman *et al.* (1988) *Proc. Natl. Acad. Sci. USA* 85:8998-9002) using the Marathon cDNA amplification kit (Clontech Laboratories Inc, Palo Alto, CA).

Additional BLAST searches are performed using the ATPT2 sequence, a sequence in the class of aromatic prenyl transferases. Additional sequences are identified in soybean libraries that are similar to the ATPT2 sequence. The additional soybean sequence demonstrates 80% identity and 91% similarity at the amino acid sequence.

Amino acid sequence alignments between ATPT2 (SEQ ID NO:2), ATPT3 (SEQ ID NO:4), ATPT4 (SEQ ID NO:6), ATPT8 (SEQ ID NO:12), and ATPT12 (SEQ ID NO:17) are performed using ClustalW (Figure 1), and the percent identity and similarities are provided in Table 1 below.

Table 1:

	ATPT2	ATPT3	ATPT4	ATPT8	ATPT12
--	-------	-------	-------	-------	--------

ATPT2	% Identity		12	13	11	15
	% similar		25	25	22	32
	% Gap		17	20	20	9
ATPT3	% Identity			12	6	22
	% similar			29	16	38
	% Gap			20	24	14
ATPT4	% Identity				9	14
	% similar				18	29
	% Gap				26	19
ATPT8	% Identity					7
	% similar					19
	% Gap					20
ATPT12	% Identity					
	% similar					
	% Gap					

Example 2: Preparation of Expression Constructs

5 A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence

10 CGCGATTAAATGGCGCGCCCTGCAGGCGGCCGCTGCAGGGCGCGCCATTAAAT (SEQ ID NO:40) was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plasmids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770, contains the pCGN7765 backbone with the napin seed specific

15 expression cassette from pCGN3223.

The cloning cassette, pCGN7787, essentially the same regulatory elements as pCGN7770, with the exception of the napin regulatory regions of pCGN7770 have been

replaced with the double CAMV 35S promoter and the tml polyadenylation and transcriptional termination region.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SmaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A series of turbo binary vectors are constructed to allow for the rapid cloning of DNA sequences into binary vectors containing transcriptional initiation regions (promoters) and transcriptional termination regions.

The plasmid pCGN8618 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:41) and 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:42) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was excised from pCGN8618 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8622.

The plasmid pCGN8619 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCC -3' (SEQ ID NO:43) and 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:44) into SalI/XhoI-digested pCGN7770. A fragment containing the napin promoter, polylinker and napin 3' region was removed from pCGN8619 by digestion with Asp718I; the fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the napin promoter was closest to the blunted Asp718I site of pCGN5139 and the napin 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8623.

The plasmid pCGN8620 was constructed by ligating oligonucleotides 5'-TCGAGGATCCGCGGCCGCAAGCTTCCTGCAGGAGCT -3' (SEQ ID NO:45) and 5'-CCTGCAGGAAGCTTGCGGCCGCGGATCC-3' (SEQ ID NO:46) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8620 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8624.

The plasmid pCGN8621 was constructed by ligating oligonucleotides 5'-TCGACCTGCAGGAAGCTTGCGGCCGCGGATCCAGCT -3' (SEQ ID NO:47) and 5'-GGATCCGCGGCCGCAAGCTTCCTGCAGG-3' (SEQ ID NO:48) into SalI/SacI-digested pCGN7787. A fragment containing the d35S promoter, polylinker and tml 3' region was removed from pCGN8621 by complete digestion with Asp718I and partial digestion with NotI. The fragment was blunt-ended by filling in the 5' overhangs with Klenow fragment then ligated into pCGN5139 that had been digested with Asp718I and HindIII and blunt-ended by filling in the 5' overhangs with Klenow fragment. A plasmid containing the insert oriented so that the d35S promoter was closest to the blunted Asp718I site of pCGN5139 and the tml 3' was closest to the blunted HindIII site was subjected to sequence analysis to confirm both the insert orientation and the integrity of cloning junctions. The resulting plasmid was designated pCGN8625.

The plasmid construct pCGN8640 is a modification of pCGN8624 described above. A 938bp PstI fragment isolated from transposon Tn7 which encodes bacterial spectinomycin and streptomycin resistance (Fling et al. (1985), *Nucleic Acids Research* 13(19):7095-7106), a determinant for E. coli and Agrobacterium selection, was blunt ended with Pfu polymerase. The blunt ended fragment was ligated into pCGN8624 that had been digested with SpeI and blunt ended with Pfu polymerase. The region containing the PstI fragment was sequenced to confirm both the insert orientation and the integrity of cloning junctions.

The spectinomycin resistance marker was introduced into pCGN8622 and pCGN8623 as follows. A 7.7 Kbp AvrII-SnaBI fragment from pCGN8640 was ligated to a 10.9 Kbp

AvrII-SnaBI fragment from pCGN8623 or pCGN8622, described above. The resulting plasmids were pCGN8641 and pCGN8643, respectively.

The plasmid pCGN8644 was constructed by ligating oligonucleotides 5'-GATCACCTGCAGGAAGCTTGCGGCCGCGGATCCAATGCA-3' (SEQ ID NO:49) and
5 5'- TTGGATCCGCGGCCGCAAGCTTCCTGCAGGT-3' (SEQ ID NO:50) into BamHI-PstI digested pCGN8640.

Synthetic oligonucleotides were designed for use in Polymerase Chain Reactions (PCR) to amplify the coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 for the preparation of expression constructs and are provided in Table 2 below.

10

Table 2:

Name	Restriction Site	Sequence	SEQ ID NO:
ATPT2	5' NotI	GGATCCGCGGCCGCACAATGGAGTC TCTGCTCTCTAGTTCT	51
ATPT2	3' SseI	GGATCCTGCAGGTCACTTCAAAAAA GGTAACAGCAAGT	52
ATPT3	5' NotI	GGATCCGCGGCCGCACAATGGCGTT TTTGGGCTCTCCCGTGTTT	53
ATPT3	3' SseI	GGATCCTGCAGGTTATTGAAAACCTT CTTCCAAGTACAACCT	54
ATPT4	5' NotI	GGATCCGCGGCCGCACAATGTGGCG AAGATCTGTTGTT	55
ATPT4	3' SseI	GGATCCTGCAGGTCACTGGAGAGTAG AAGGAAGGAGCT	56
ATPT8	5' NotI	GGATCCGCGGCCGCACAATGGTACT TGCCGAGGTTCCAAAGCTTGCTCT	57
ATPT8	3' SseI	GGATCCTGCAGGTCACTTGTCTCTG GTGATGACTCTAT	58
ATPT12	5' NotI	GGATCCGCGGCCGCACAATGACTTC GATTCTCAACACT	59
ATPT12	3' SseI	GGATCCTGCAGGTCACTGTTGCGAT GCTAATGCCGT	60

The coding sequences of ATPT2, ATPT3, ATPT4, ATPT8, and ATPT12 were all amplified using the respective PCR primers shown in Table 2 above and cloned into the TopoTA
15 vector (Invitrogen). Constructs containing the respective prenyltransferase sequences were digested with NotI and Sse8387I and cloned into the turbobinary vectors described above.

The sequence encoding ATPT2 prenyltransferase was cloned in the sense orientation into pCGN8640 to produce the plant transformation construct pCGN10800 (Figure 2). The ATPT2 sequence is under control of the 35S promoter.

The ATPT2 sequence was also cloned in the antisense orientation into the construct pCGN8641 to create pCGN10801 (Figure 3). This construct provides for the antisense expression of the ATPT2 sequence from the napin promoter.

5 The ATPT2 coding sequence was also cloned in the antisense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10802

The ATPT2 coding sequence was also cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10803 (Figure 4).

10 The ATPT4 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10806 (Figure 5). The ATPT2 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10807 (Figure 6). The ATPT3 coding sequence was cloned into the vector pCGN864 to create the plant transformation construct pCGN10808 (Figure 7). The ATPT3 coding sequence was cloned in the sense orientation into the vector pCGN8640 to create the plant transformation construct pCGN10809 (Figure 8). The ATPT3 coding sequence was cloned in the antisense orientation into the vector
15 pCGN8641 to create the plant transformation construct pCGN10810 (Figure 9). The ATPT3 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10811 (Figure 10). The ATPT3 coding sequence was cloned into the vector pCGN8640 to create the plant transformation construct pCGN10812 (Figure 11). The ATPT4 coding sequence was cloned into the vector pCGN8640 to create the plant transformation
20 construct pCGN10813 (Figure 12). The ATPT4 coding sequence was cloned into the vector pCGN8643 to create the plant transformation construct pCGN10814 (Figure 13). The ATPT4 coding sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10815 (Figure 14). The ATPT4 coding sequence was cloned in the antisense orientation into the vector pCGN8644 to create the plant transformation construct pCGN10816
25 (Figure 15). The ATPT2 coding sequence was cloned into the vector pCGN???? to create the plant transformation construct pCGN10817 (Figure 16). The ATPT8 coding sequence was cloned in the sense orientation into the vector pCGN8643 to create the plant transformation construct pCGN10819 (Figure 17). The ATPT12 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct pCGN10824 (Figure 18). The ATPT12 coding
30 sequence was cloned into the vector pCGN8641 to create the plant transformation construct pCGN10825 (Figure 19). The ATPT8 coding sequence was cloned into the vector pCGN8644 to create the plant transformation construct pCGN10826 (Figure 20).

Example 3: Plant Transformation

5 Transgenic *Brassica* plants are obtained by *Agrobacterium*-mediated transformation as described by Radke *et al.* (*Theor. Appl. Genet.* (1988) 75:685-694; *Plant Cell Reports* (1992) 11:499-505). Transgenic *Arabidopsis thaliana* plants may be obtained by *Agrobacterium*-mediated transformation as described by Valverkens *et al.*, (*Proc. Nat. Acad. Sci.* (1988) 85:5536-5540), or as described by Bent *et al.* ((1994), *Science* 265:1856-1860), or
 10 Bechtold *et al.* ((1993), *C.R.Acad.Sci, Life Sciences* 316:1194-1199). Other plant species may be similarly transformed using related techniques.

Alternatively, microprojectile bombardment methods, such as described by Klein *et al.* (*Bio/Technology* 10:286-291) may also be used to obtain nuclear transformed plants.

15

Example 4: Identification of Additional Prenyltransferases

A PSI-Blast profile generated using the *E. coli* *ubiA* (genbank accession 1790473) sequence was used to analyze the *Synechocystis* genome. This analysis identified 5 open
 20 reading frames (ORFs) in the *Synechocystis* genome that were potentially prenyltransferases; *slr0926* (annotated as *ubiA* (4-hydroxybenzoate-octaprenyl transferase, SEQ ID NO:32), *slr1899* (annotated as *ctaB* (cytochrome c oxidase folding protein, SEQ ID NO:33), *slr0056* (annotated as *g4* (chlorophyll synthase 33 kd subunit, SEQ ID NO:34), *slr1518* (annotated as *menaA* (menaquinone biosynthesis protein, SEQ ID NO:35), and *slr1736* (annotated as a
 25 hypothetical protein of unknown function (SEQ ID NO:36).

To determine the functionality of these ORFs and their involvement, if any, in the biosynthesis of Tocopherols, knockouts constructs were made to disrupt the ORF identified in *Synechocystis*.

Synthetic oligos were designed to amplify regions from the 5' (5'-
 30 TAATGTGTACATTGTCGGCCTC (17365') (SEQ ID NO:61) and 5'-
 GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCCACAATTCCCCGCA
 CCGTC (1736kanpr1)) (SEQ ID NO:62) and 3' (5'-AGGCTAATAAGCACAAATGGGA
 (17363') (SEQ ID NO:63) and 5'-

GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGC

GGAATTGGTTTAGGTTATCCC (1736kanpr2)) (SEQ ID NO:64) ends of the slr1736 ORF.

The 1736kanpr1 and 1736kanpr2 oligos contained 20 bp of homology to the slr1736 ORF with an additional 40 bp of sequence homology to the ends of the kanamycin resistance

5 cassette. Separate PCR steps were completed with these oligos and the products were gel purified and combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The combined fragments were allowed to assemble without oligos under the following conditions: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min plus 5 seconds per cycle for 40 cycles using pfu
10 polymerase in 100ul reaction volume (Zhao, H and Arnold (1997) *Nucleic Acids Res.* 25(6):1307-1308). One microliter or five microliters of this assembly reaction was then amplified using 5' and 3' oligos nested within the ends of the ORF fragment, so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be
15 knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21681 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The *ubiA* 5' sequence was amplified using the primers 5'-

20 GGATCCATGGTT GCCCAAACCCCATC (SEQ ID NO:65) and 5'-

GCAATGTAACATCAGAGA TTTTGAGACACAACG

TGGCTTTGGGTAAGCAACAATGACCGGC (SEQ ID NO:66). The 3' region was amplified using the synthetic oligonucleotide primers 5'-

GAATTCTCAAAGCCAGCCCAGTAAC (SEQ ID NO:67) and 5'-GGTATGAGTC

25 AGCAACACCTTCTTCACGAGGCAGACCTCAGCGGGTGCAGAAAAGGGTTTTCCC

(SEQ ID NO:68). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested

within the ends of the ORF fragment (5'-CCAGTGGTTTAGGCTGTGTGGTC (SEQ ID

30 NO:69) and 5'-CTGAGTTGGATGTATTGGATC (SEQ ID NO:70)), so that the resulting

product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out.

This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21682 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The sl11899 5' sequence was amplified using the primers 5'-
 5 GGATCCATGGTACTT CGACAAAAATCC (SEQ ID NO:71) and 5'-
 GCAATGTAACATCAGAG
 ATTTTGAGACACAACGTGGCTTTGCTAGGCAACCGCTTAGTAC (SEQ ID NO:72).

The 3' region was amplified using the synthetic oligonucleotide primers 5'-

10 GAATTCTTAACCCAACAGTAAAGTTCCC (SEQ ID NO:73) and 5'-
 GGTATGAGTCAGC
 AACACCTTCTTCACGAGGCAGACCTCAGCGCCGGCATTGTCTTTTACATG (SEQ ID
 NO:74). The amplification products were combined with the kanamycin resistance gene from
 puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector
 15 backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the
 ends of the ORF fragment (5'- GGAACCCTTGCGCCGCTTC (SEQ ID NO:75)
 and 5'- GTATGCCCAACTGGTGCAGAGG (SEQ ID NO:76)), so that the resulting product
 contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the
 kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out.
 20 This PCR product was then cloned into the vector pGemT easy (Promega) to create the
 construct pMON21679 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The slr0056 5' sequence was amplified using the primers 5'-

25 GGATCCATGTCTGACACACAAAATACCG (SEQ ID NO:77) and 5'-
 GCAATGTAACATCAGAGATTTTGAGACACAACGTGGCTTTCGCCAATACCAGCCA
 CCAACAG (SEQ ID NO:78). The 3' region was amplified using the synthetic
 oligonucleotide primers 5'- GAATTCTCAAAT CCCCGCATGGCCTAG (SEQ ID NO:79)
 and 5'-
 30 GGTATGAGTCAGCAACACCTTCTTCACGAGGCAGACCTCAGCGGCCTACGGCTTG
 GACGTGTGGG (SEQ ID NO:80). The amplification products were combined with the
 kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and
 gel purified away from the vector backbone. The annealed fragment was amplified using 5'

and 3' oligos nested within the ends of the ORF fragment (5'-CACTTGGATTCCCCTGATCTG (SEQ ID NO:81) and 5'-GCAATACCCGCTTGGAAAACG (SEQ ID NO:82)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21677 and used for *Synechocystis* transformation.

Primers were also synthesized for the preparation of *Synechocystis* knockout constructs for the other sequences using the same method as described above, with the following primers. The slr1518 5' sequence was amplified using the primers 5'-GGATCCATGACCGAATCTTCGCCCCTAGC (SEQ ID NO:83) and 5'-GCAATGTAACATCAGAGATTTTGA GACACAACGTGGC TTTCAATCCTAGGTAGCCGAGGCG (SEQ ID NO:84). The 3' region was amplified using the synthetic oligonucleotide primers 5'-GAATTCTTAGCCCAGGCC AGCCCAGCC (SEQ ID NO:85) and 5'-GGTATGAGTCAGCAACACCTTCTTCACGA GGCAGACCTCAGCGGGGAATTGATTTGTTTAATTACC (SEQ ID NO:86). The amplification products were combined with the kanamycin resistance gene from puc4K (Pharmacia) that had been digested with *HincII* and gel purified away from the vector backbone. The annealed fragment was amplified using 5' and 3' oligos nested within the ends of the ORF fragment (5'-GCGATCGCCATTATCGCTTGG (SEQ ID NO:87) and 5'-GCAGACTGGCAATTATCAGTAACG (SEQ ID NO:88)), so that the resulting product contained 100-200 bp of the 5' end of the *Synechocystis* gene to be knocked out, the kanamycin resistance cassette, and 100-200 bp of the 3' end of the gene to be knocked out. This PCR product was then cloned into the vector pGemT easy (Promega) to create the construct pMON21680 and used for *Synechocystis* transformation.

B. Transformation of *Synechocystis*

Cells of *Synechocystis* 6803 were grown to a density of approximately 2×10^8 cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium (ATCC Medium 616) at a density of 1×10^9 cells per ml and used immediately for transformation. One-hundred microliters of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8 and allowed to grow for 12-18

hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing

5 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

10 PCR analysis of the *Synechocystis* isolates for slr1736 and sl1899 showed complete segregation of the mutant genome, meaning no copies of the wild type genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of these same isolates showed that the sl1899 strain had no detectable reduction in tocopherol levels. However, the strain carrying the knockout for slr1736 produced no detectable levels of tocopherol.

15 The amino acid sequences for the *Synechocystis* knockouts are compared using ClustalW, and are provided in Table 3 below. Provided are the percent identities, percent similarity, and the percent gap. The alignment of the sequences is provided in Figure 21.

Table 3:

	Slr1736	slr0926	sl1899	slr0056	slr1518
slr1736 %identity	14	12	18	11	
%similar	29	30	34	26	
%gap	8	7	10	5	
slr0926 %identity		20	19	14	
%similar		39	32	28	
%gap		7	9	4	
sl1899 %identity			17	13	
%similar			29	29	
%gap			12	9	
slr0056 %identity				15	
%similar				31	
%gap				8	

slr1518 %identity	
%similar	
%gap	

Amino acid sequence comparisons are performed using various *Arabidopsis* prenyltransferase sequences and the *Synechocystis* sequences. The comparisons are presented in Table 4 below. Provided are the percent identities, percent similarity, and the percent gap.

- 5 The alignment of the sequences is provided in Figure 22.

Table 4:

	ATPT2	slr1736	ATPT3	slr0926	ATPT4	sll1899	ATPT12	slr0056	ATPT8	slr1518
ATPT2		29	9	9	8	8	12	9	7	9
		46	23	21	20	20	28	23	21	20
		27	13	28	23	29	11	24	25	24
slr1736			9	13	8	12	13	15	8	10
			19	28	19	28	26	33	21	26
			34	12	34	15	26	10	12	10
ATPT3				23	11	14	13	10	5	11
				36	26	26	26	21	14	22
				29	21	31	16	30	30	30
					12	20	17	20	11	14
slr0926					24	37	28	33	24	29
					33	12	25	10	11	9
						18	11	8	6	7
ATPT4						33	23	18	16	19
						28	19	32	32	33
							13	17	10	12
sll1899							24	30	23	26
							27	13	10	11
								52	8	11
ATPT1								66	19	26
2								18	25	23

	9	13
slr0056	23	32
	10	8
		7
ATPT8		23
		7
slr1518		

4B. Preparation of the slr1737 Knockout

The *Synechocystis* sp. 6803 slr1737 knockout was constructed by the following method. The GPS™-1 Genome Priming System (New England Biolabs) was used to insert, by a Tn7 Transposase system, a Kanamycin resistance cassette into *slr1737*. A plasmid from a *Synechocystis* genomic library clone containing 652 base pairs of the targeted orf (*Synechocystis* genome base pairs 1324051 – 1324703; the predicted orf base pairs 1323672 – 1324763, as annotated by Cyanobase) was used as target DNA. The reaction was performed according to the manufacturers protocol. The reaction mixture was then transformed into *E. coli* DH10B electrocompetant cells and plated. Colonies from this transformation were then screened for transposon insertions into the target sequence by amplifying with M13 Forward and Reverse Universal primers, yielding a product of 652 base pairs plus ~1700 base pairs, the size of the transposon kanamycin cassette, for a total fragment size of ~2300 base pairs. After this determination, it was then necessary to determine the approximate location of the insertion within the targeted orf, as 100 base pairs of orf sequence was estimated as necessary for efficient homologous recombination in *Synechocystis*. This was accomplished through amplification reactions using either of the primers to the ends of the transposon, Primer S (5' end) or N (3' end), in combination with either a M13 Forward or Reverse primer. That is, four different primer combinations were used to map each potential knockout construct: Primer S – M13 Forward, Primer S – M13 Reverse, Primer N – M13 Forward, Primer N – M13 Reverse. The construct used to transform *Synechocystis* and knockout slr1737 was determined to consist of a approximately

150 base pairs of slr1737 sequence on the 5' side of the transposon insertion and approximately 500 base pairs on the 3' side, with the transcription of the orf and kanamycin cassette in the same direction. The nucleic acid sequence of slr1737 is provided in SEQ ID NO:38 the deduced amino acid sequence is provided in SEQ ID NO:39.

5 Cells of *Synechocystis* 6803 were grown to a density of $\sim 2 \times 10^8$ cells per ml and harvested by centrifugation. The cell pellet was re-suspended in fresh BG-11 medium at a density of 1×10^9 cells per ml and used immediately for transformation. 100 ul of these cells were mixed with 5 ul of mini prep DNA and incubated with light at 30C for 4 hours. This mixture was then plated onto nylon filters resting on BG-11 agar supplemented with TES pH8
10 and allowed to grow for 12-18 hours. The filters were then transferred to BG-11 agar + TES + 5ug/ml kanamycin and allowed to grow until colonies appeared within 7-10 days (Packer and Glazer, 1988). Colonies were then picked into BG-11 liquid media containing 5 ug/ml kanamycin and allowed to grow for 5 days. These cells were then transferred to Bg-11 media containing 10ug/ml kanamycin and allowed to grow for 5 days and then transferred to Bg-11
15 + kanamycin at 25ug/ml and allowed to grow for 5 days. Cells were then harvested for PCR analysis to determine the presence of a disrupted ORF and also for HPLC analysis to determine if the disruption had any effect on tocopherol levels.

 PCR analysis of the *Synechocystis* isolates, using primers to the ends of the *slr1737* orf, showed complete segregation of the mutant genome, meaning no copies of the wild type
20 genome could be detected in these strains. This suggests that function of the native gene is not essential for cell function. HPLC analysis of the strain carrying the knockout for *slr1737* produced no detectable levels of tocopherol.

4C. Phytyl Prenyltransferase Enzyme Assays

25 [³H] Homogentisic acid in 0.1% H₃PO₄ (specific radioactivity 40 Ci/mmol). Phytyl pyrophosphate was synthesized as described by Joo, *et al.* (1973) *Can J. Biochem.* 51:1527. 2-methyl-6-phytylquinol and 2,3-dimethyl-5-phytylquinol were synthesized as described by Soll, *et al.* (1980) *Phytochemistry* 19:215. Homogentisic acid, α , β , δ , and γ -tocopherol, and tocol, were purchased commercially.

30 The wild-type strain of *Synechocystis* sp. PCC 6803 was grown in BG11 medium with bubbling air at 30°C under 50 $\mu\text{E} \cdot \text{m}^{-2} \cdot \text{s}^{-1}$ fluorescent light, and 70% relative humidity. The growth medium of slr1736 knock-out (potential PPT) strain of this organism was

supplemented with 25 $\mu\text{g mL}^{-1}$ kanamycin. Cells were collected from 0.25 to 1 liter culture by centrifugation at 5000 g for 10 min and stored at -80°C .

Total membranes were isolated according to Zak's procedures with some modifications (Zak, *et al.* (1999) *Eur J. Biochem* 261:311). Cells were broken on a French press. Before the French press treatment, the cells were incubated for 1 hour with lysozyme (0.5%, w/v) at 30°C in a medium containing 7 mM EDTA, 5 mM NaCl and 10 mM Hepes-NaOH, pH 7.4. The spheroplasts were collected by centrifugation at 5000 g for 10 min and resuspended at 0.1 - 0.5 mg chlorophyll $\cdot\text{mL}^{-1}$ in 20 mM potassium phosphate buffer, pH 7.8. Proper amount of protease inhibitor cocktail and DNAase I from Boehringer Mannheim were added to the solution. French press treatments were performed two to three times at 100 MPa. After breakage, the cell suspension was centrifuged for 10 min at 5000 g to pellet unbroken cells, and this was followed by centrifugation at 100 000 g for 1 hour to collect total membranes. The final pellet was resuspended in a buffer containing 50 mM Tris-HCL and 4 mM MgCl_2 .

Chloroplast pellets were isolated from 250 g of spinach leaves obtained from local markets. Devined leaf sections were cut into grinding buffer (2 l /250 g leaves) containing 2 mM EDTA, 1 mM MgCl_2 , 1 mM MnCl_2 , 0.33 M sorbitol, 0.1% ascorbic acid, and 50 mM Hepes at pH 7.5. The leaves were homogenized for 3 sec three times in a 1-L blender, and filtered through 4 layers of microcloth. The supernatant was then centrifuged at 5000 g for 6 min. The chloroplast pellets were resuspended in small amount of grinding buffer (Douce, *et al* Methods in Chloroplast Molecular Biology, 239 (1982)

Chloroplasts in pellets can be broken in three ways. Chloroplast pellets were first aliquoted in 1 mg of chlorophyll per tube, centrifuged at 6000 rpm for 2 min in microcentrifuge, and grinding buffer was removed. Two hundred microliters of Triton X-100 buffer (0.1% Triton X-100, 50 mM Tris-HCl pH 7.6 and 4 mM MgCl_2) or swelling buffer (10 mM Tris pH 7.6 and 4 mM MgCl_2) was added to each tube and incubated for $\frac{1}{2}$ hour at 4°C . Then the broken chloroplast pellets were used for the assay immediately. In addition, broken chloroplasts can also be obtained by freezing in liquid nitrogen and stored at -80°C for $\frac{1}{2}$ hour, then used for the assay.

In some cases chloroplast pellets were further purified with 40%/ 80% percoll gradient to obtain intact chloroplasts. The intact chloroplasts were broken with swelling buffer, then either used for assay or further purified for envelope membranes with 20.5%/ 31.8% sucrose density gradient (Sol, *et al* (1980) *supra*). The membrane fractions were centrifuged at 100 000 g for 40 min and resuspended in 50 mM Tris-HCl pH 7.6, 4 mM MgCl_2 .

Various amounts of [^3H]HGA, 40 to 60 μM unlabelled HGA with specific activity in the range of 0.16 to 4 Ci/mmol were mixed with a proper amount of 1M Tris-NaOH pH 10 to adjust pH to 7.6. HGA was reduced for 4 min with a trace amount of solid NaBH_4 . In addition to HGA, standard incubation mixture (final vol 1 mL) contained 50 mM Tris-HCl, pH 7.6, 3-5 mM MgCl_2 , and 100 μM phytyl pyrophosphate. The reaction was initiated by addition of *Synechocystis* total membranes, spinach chloroplast pellets, spinach broken chloroplasts, or spinach envelope membranes. The enzyme reaction was carried out for 2 hour at 23°C or 30°C in the dark or light. The reaction is stopped by freezing with liquid nitrogen, and stored at -80°C or directly by extraction.

A constant amount of tocol was added to each assay mixture and reaction products were extracted with a 2 mL mixture of chloroform/methanol (1:2, v/v) to give a monophasic solution. NaCl solution (2 mL; 0.9%) was added with vigorous shaking. This extraction procedure was repeated three times. The organic layer containing the prenylquinones was filtered through a 20 μm filter, evaporated under N_2 , and then resuspended in 100 μL of ethanol.

The samples were mainly analyzed by Normal-Phase HPLC method (Isocratic 90% Hexane and 10% Methyl-t-butyl ether), and use a Zorbax silica column, 4.6 x 250 mm. The samples were also analyzed by Reversed-Phase HPLC method (Isocratic 0.1% H_3PO_4 in MeOH), and use a Vydac 201HS54 C18 column; 4.6 x 250 mm coupled with an All-tech C18 guard column. The amount of products were calculated based on the substrate specific radioactivity, and adjusted according to the % recovery based on the amount of internal standard.

The amount of chlorophyll was determined as described in Arnon (1949) *Plant Physiol.* 24:1. Amount of protein was determined by the Bradford method using gamma globulin as a standard (Bradford, (1976) *Anal. Biochem.* 72:248)

Results of the assay demonstrate that 2-Methyl-6-Phytylplastoquinone is produced in the *Synechocystis* slr1736 knockout preparations. The results of the phytyl prenyltransferase enzyme activity assay for the slr1736 knock out are presented in Figure 23.

4D. Complementation of the slr1736 knockout with ATPT2

In order to determine whether ATPT2 could complement the knockout of slr1736 in *Synechocystis* 6803 a plasmid was constructed to express the ATPT2 sequence from the TAC promoter. A vector, plasmid psl1211, was obtained from the lab of Dr. Himadri Pakrasi of

Washington University, and is based on the plasmid RSF1010 which is a broad host range plasmid (Ng W.-O., Zentella R., Wang, Y., Taylor J-S. A., Pakrasi, H.B. 2000. *phrA*, the major photoreactivating factor in the cyanobacterium *Synechocystis* sp. strain PCC 6803 codes for a cyclobutane pyrimidine dimer specific DNA photolyase. *Arch. Microbiol.* (in press)). The ATPT2 gene was isolated from the vector pCGN10817 by PCR using the following primers. ATPT2_{nco.pr} 5'-CCATGGATTCTGAGTAAAGTTGTCTCG (SEQ ID NO:89); ATPT2_{ri.pr} 5'-GAATTCACCTTCAAAAAGGTAACAG (SEQ ID NO:90). These primers will remove approximately 112 BP from the 5' end of the ATPT2 sequence, which is thought to be the chloroplast transit peptide. These primers will also add an NcoI site at the 5' end and an EcoRI site at the 3' end which can be used for sub-cloning into subsequent vectors. The PCR product from using these primers and pCGN10817 was ligated into pGEM T easy and the resulting vector pMON21689 was confirmed by sequencing using the m13forward and m13reverse primers. The NcoI/EcoRI fragment from pMON21689 was then ligated with the EagI/EcoRI and EagI/NcoI fragments from psl1211 resulting in pMON21690. The plasmid pMON21690 was introduced into the slr1736 *Synechocystis* 6803 KO strain via conjugation. Cells of sl906 (a helper strain) and DH10B cells containing pMON21690 were grown to log phase (O.D. 600= 0.4) and 1 ml was harvested by centrifugation. The cell pellets were washed twice with a sterile BG-11 solution and resuspended in 200 ul of BG-11. The following was mixed in a sterile eppendorf tube: 50 ul SL906, 50 ul DH10B cells containing pMON21690, and 100 ul of a fresh culture of the slr1736 *Synechocystis* 6803 KO strain (O.D. 730 = 0.2-0.4). The cell mixture was immediately transferred to a nitrocellulose filter resting on BG-11 and incubated for 24 hours at 30C and 2500 LUX(50 ue) of light. The filter was then transferred to BG-11 supplemented with 10ug/ml Gentamycin and incubated as above for ~5 days. When colonies appeared, they were picked and grown up in liquid BG-11 + Gentamycin 10 ug/ml. (Elhai, J. and Wolk, P. 1988. Conjugal transfer of DNA to Cyanobacteria. *Methods in Enzymology* 167, 747-54) The liquid cultures were then assayed for tocopherols by harvesting 1ml of culture by centrifugation, extracting with ethanol/pyrogallol, and HPLC separation. The slr1736 *Synechocystis* 6803 KO strain, did not contain any detectable tocopherols, while the slr1736 *Synechocystis* 6803 KO strain transformed with pmon21690 contained detectable alpha tocopherol. A *Synechocystis* 6803 strain transformed with psl1211(vector control) produced alpha tocopherol as well.

Example 5: Transgenic Plant Analysis

Arabidopsis plants transformed with constructs for the sense or antisense expression
 5 of the ATPT proteins were analyzed by High Pressure Liquid Chromatography (HPLC) for
 altered levels of total tocopherols, as well as altered levels of specific tocopherols (alpha,
 beta, gamma, and delta tocopherol).

Extracts of leaves and seeds were prepared for HPLC as follows. For seed extracts,
 10 mg of seed was added to 1 g of microbeads (Biospec) in a sterile microfuge tube to which
 10 500 ul 1% pyrogallol (Sigma Chem)/ethanol was added. The mixture was shaken for 3
 minutes in a mini Beadbeater (Biospec) on "fast" speed. The extract was filtered through a
 0.2 um filter into an autosampler tube. The filtered extracts were then used in HPLC analysis
 described below.

Leaf extracts were prepared by mixing 30-50 mg of leaf tissue with 1 g microbeads
 15 and freezing in liquid nitrogen until extraction. For extraction, 500 ul 1% pyrogallol in
 ethanol was added to the leaf/bead mixture and shaken for 1 minute on a Beadbeater
 (Biospec) on "fast" speed. The resulting mixture was centrifuged for 4 minutes at 14,000 rpm
 and filtered as described above prior to HPLC analysis.

HPLC was performed on a Zorbax silica HPLC column (4.6 mm X 250 mm) with a
 20 fluorescent detection, an excitation at 290 nm, an emission at 336 nm, and bandpass and slits.
 Solvent A was hexane and solvent B was methyl-t-butyl ether. The injection volume was 20
 ul, the flow rate was 1.5 ml/min, the run time was 12 min (40°C) using the gradient (Table 5):

Table 5:

<u>Time</u>	<u>Solvent A</u>	<u>Solvent B</u>
0 min.	90%	10%
10 min.	90%	10%
11 min.	25%	75%
12 min.	90%	10%

Tocopherol standards in 1% pyrogallol/ ethanol were also run for comparison (alpha
 tocopherol, gamma tocopherol, beta tocopherol, delta tocopherol, and tocopherol (tocol) (all
 from Matreya).

Standard curves for alpha, beta, delta, and gamma tocopherol were calculated using
 35 Chemstation software. The absolute amount of component x is: Absolute amount of x=

$\text{Response}_x \times \text{RF}_x \times \text{dilution factor}$ where Response_x is the area of peak x , RF_x is the response factor for component x ($\text{Amount}_x/\text{Response}_x$) and the dilution factor is 500 μl . The ng/mg tissue is found by: total ng component/ mg plant tissue.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10822 for the expression of ATAT2 from the napin promoter are provided in Figure 24.

HPLC analysis results of *Arabidopsis* seed tissue expressing the ATAT2 sequence from the napin promoter (pMON10822) demonstrates an increased level of tocopherols in the seed. Total tocopherol levels are increased as much as 50 to 60% over the total tocopherol levels of non-transformed (wild-type) *Arabidopsis* plants (Figure 24).

Furthermore, increases of particular tocopherols are also increased in transgenic *Arabidopsis* plants expressing the ATAT2 nucleic acid sequence from the napin promoter. Levels of delta tocopherol in these lines are increased greater than 3 fold over the delta tocopherol levels obtained from the seeds of wild type *Arabidopsis* lines. Levels of gamma tocopherol in transgenic *Arabidopsis* lines expressing the ATAT2 nucleic acid sequence are increased as much as about 60% over the levels obtained in the seeds of non-transgenic control lines. Furthermore, levels of alpha tocopherol are increased as much as 3 fold over those obtained from non-transgenic control lines.

Results of the HPLC analysis of seed extracts of transgenic *Arabidopsis* lines containing pMON10803 for the expression of ATAT2 from the enhanced 35S promoter are provided in Figure 25.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claim.

Claims

What is Claimed is:

1. An isolated nucleic acid sequence encoding a prenyltransferase.
- 5 2. An isolated nucleic acid sequence according to Claim 1, wherein said prenyltransferase is selected from the group consisting of straight chain prenyltransferase and aromatic prenyltransferase.
3. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is isolated from a eukaryotic cell source.
4. An isolated DNA sequence according to Claim 3, wherein said eukaryotic cell source is
10 selected from the group consisting of mammalian, nematode, fungal, and plant cells.
5. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from *Arabidopsis*.
6. The DNA encoding sequence of Claim 5 wherein said prenyltransferase protein is encoded by a sequence selected from the group consisting of the sequences of Figure 1.
- 15 7. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from corn.
8. The DNA encoding sequence of Claim 7 wherein said prenyltransferase protein is encoded by a sequence which includes the EST of the sequences of Figure 3.
9. The DNA encoding sequence of Claim 4 wherein said prenyltransferase protein is from
20 soybean.
10. The DNA encoding sequence of Claim 9 wherein said prenyltransferase protein is encoded by a sequence which includes the ESTs of the group consisting of the sequences of Figure 2 and Figure 9.
11. An isolated DNA sequence according to Claim 1, wherein said nucleic acid sequence is
25 isolated from a prokaryotic cell source.
12. An isolated DNA sequence according to Claim 11, wherein said prokaryotic source is *Synechocystis*.
13. A nucleic acid construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding prenyltransferase, and a
30 transcriptional termination region.
14. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding prenyltransferase is obtained from an organism selected from the group consisting of a eukaryotic organism and a prokaryotic organism.

15. A nucleic acid construct according to Claim 14, wherein said nucleic acid sequence encoding prenyltransferase is obtained from a plant source.

16. A nucleic acid construct according to Claim 15, wherein said nucleic acid sequence encoding prenyltransferase is obtained from a source selected from the group consisting of
5 *Arabidopsis*, soybean and corn.

17. A nucleic acid construct according to Claim 13, wherein said nucleic acid sequence encoding prenyltransferase is obtained from *Synechocystis*.

18. A plant cell comprising the construct of Claim 13.

19. A method for the alteration of the tocopherol content in a host cell, comprising;
10 transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a host cell, a nucleic acid sequence encoding prenyltransferase, and a transcriptional termination region.

20. The method according to Claim 19, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

15 21. The method according to Claim 20, wherein said prokaryotic cell is *Synechocystis*.

22. The method according to Claim 20, wherein said eukaryotic cell is a plant cell.

23. The method according to Claim 22, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

24. A method for producing a tocopherol compound of interest in a host cell, said method
20 comprising obtaining a transformed host cell, said host cell having and expressing in its genome: a construct having a DNA sequence encoding a prenyltransferase operably linked to a transcriptional initiation region functional in a host cell,
wherein said prenyltransferase is involved in the synthesis of tocopherols.

25. The method according to Claim 24, wherein said host cell is selected from the group
25 consisting of a prokaryotic cell and a eukaryotic cell.

26. The method according to Claim 25, wherein said prokaryotic cell is *Synechocystis*.

27. The method according to Claim 24, wherein said eukaryotic cell is a plant cell.

28. The method according to Claim 27, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

30 29. A method for increasing the biosynthetic flux in cell from a host cell toward tocopherol production, said method comprising transforming said host cell with a construct comprising as operably linked components, a transcriptional initiation region functional in a

host cell, a DNA encoding a prenyltransferase involved in the synthesis of tocopherols, and a transcriptional termination region.

30. The method according to Claim 29, wherein said host cell is selected from the group consisting of a prokaryotic cell and a eukaryotic cell.

5 31. The method according to Claim 30, wherein said prokaryotic cell is *Synechocystis*.

32. The method according to Claim 30, wherein said eukaryotic cell is a plant cell.

33. The method according to Claim 32, wherein said plant cell is obtained from a plant selected from the group consisting of *Arabidopsis*, soybean, and corn.

10

ATPT2	-----MES-----LSSSELVAAGG-----FCVKKON-----LKLHSLSLHIVLRCDSSKVAKPA-----RR-----NNLVRP-----DGQG : 59	*	20	*	40	*	60	*	80	*
ATPT3	MAFFGLSRVSRRLHKEVSVPSPS SALLQSHKHLSPVTHYTNPTKCYPMNDNYQVMSKREHLPQVGFVGMVYRLICGMSSSS : 90									
ATPT4	-----MWRRS-----VWYRFSRRLSVSSLPNPRIPVREL-----CAVNSPSPVPTETAKGITGV-----KSD-----ANRVFA-----ATA : 69									
ATPT8	-----MVLAE-----PKLA-----AAEYFFKR-----GVQKQF-----SHILLMATA-----LN-----VRPE-----ALIG : 48									
ATPT12	-----MTS-----LHNVETLHSSRVTSVDRVGVLSLRN-----SDSVEFRRR-----SGFSLLIESPK-----HVV-----VRAAE-----RDD : 64									
ATPT2	SSLLLY-----KHKSRE-----TATAGPEAFDSNSKQKSRDS-----AFYRFSR-----HIVLGTVLSILSVSELAWEKVSDISPLFTGILEA : 141		100	*	120	*	140	*	160	*
ATPT3	SVLEGGKKDDKEKSDGVVKKASWDLYLPEEVRGYAKLARLKKPGTWLLAWCVNSALAADPSLPSF-----K-----YMAFGCG : 171									
ATPT4	AATATAT-----TG-EISS-----VALAGHGHYARCVMELSK-AKLSMLVATS-----GTGVGT-GNAAISFPGL-----C-----YTCA : 137									
ATPT8	ESTDIVT-----SELVRQRGIEITEMIHVASLLHDDVL-DDATRRGVGS-----LNIVGKMSVLAGD-----LLS-----RACG : 117									
ATPT12	KVKSQTE-----DKAPAGGSSINQLLCKG-ASQETNKKIRLQITKPN-TWP-----ELANGVVCAGAAASGNEHWTPED-----VAKSILCM : 140									
ATPT2	VVAALMNIYIVGLNQSSEVERKVNKP-----YLPLASGEYSVNTGIAIASPSSMS-FWLGWVGSGSPLFWA-----LFSFGLGTASINL : 224		200	*	220	*	240	*	260	*
ATPT3	AL-----LRGAGC-----INDLEODTKVDRTKLHPIASCLTPFQGGICGLOLLH-LGILLQNNKS-----RMLGASLILLVSY : 248									
ATPT4	GT-----MTAASANSINQIFESNBSKMKRTMLRPLPSEHSPVPHAVATAGASSACLASKTMLAAG-----LASANVLYAVATP : 219									
ATPT8	AT-----AALKNTEVALLATAHEHLYTGET-----MEITSSPEQRYSMDYMKQTVYKT-----ASLSNSCK-----AAVETGQTAEVAV : 190									
ATPT12	MSGPCITGYTQITINDWYDRDEINAINEP-----YRHPSCATSEPEVITQVWVLLGCG-LGIAGITVWAGHTTPTVLYALGCGSLLSVYSA : 227									
ATPT2	BLRWRRPALVAMCIHVRRAI-----AFVLIHQTHFGRPILFTRPIFAAFMSFFS-VVLAHFKDIPDEG-----D-----KI : 299		280	*	300	*	320	*	340	*
ATPT3	B-----EMKRFTRPQDFEGIT-----INWGALFMT-----EMKCSRAPSIHVP-----LYLSCQCWTLVYDTHAHQDKED-----D-----VK : 314									
ATPT4	LKQLHPINTWVGVV-----GAPPLLEA-----AASQOBYNSMHPHMAHCRNDYAAAGYKMLSFPDPSGKRIAA : 300									
ATPT8	LAFEYGRILGAFQI-----DDMDFTS-----ASLCKGSLDIRHGVIAPILFAMEEFPQIREVVDDQEK-----DP-----RN : 259									
ATPT12	BPTKLKONGMVGNEFA-EG-ASYHSLPWAAGQ-----AFGITHRPDVVVI-----MLLYSHAG-LGIATVANDFKSVEG-----D-----RA : 294									
ATPT2	FGIR-----FSVTLIQ-----KRVFTG-----VTLQVAYVAILVGAATSPFTTSK-----LSVGHVLAATTLWARAKSVLSSSKTEITSCM : 375		380	*	400	*	420	*	440	*
ATPT3	VGINS-----TALFEGD-----NKLMTGFGASIGFTALSGSADLGWOYKAS-----LAAASQGGVQIGTADLSSGALCSRKFVSNKWA : 392									
ATPT4	VALRNCFYMIPGHIAYDWGLSSMFLSTELTHTATATPSFYDRTHHKARKMFHSMFLPVMGSLLLHRSVNDNQQLVEEAGL : 390									
ATPT8	VDIAL-----EYLGKSK-----GLO-----RARELMEHNLAAAGISIPET-----DNEDVKRSRRALIDLTHRVITRNK----- : 321									
ATPT12	LGLOS-----LPLAFCT-----ETAKMIG-VGADITQISVAGLLASGKPYALA-LVALIPQIVQFYFLKDPVKYDVVKYQASAQPH : 373									
ATPT2	MFWLWLFYAE-----YLLLPFLK----- : 393		460	*	480	*				
ATPT3	GAIHFSGVWLG-----RSFG----- : 407									
ATPT4	TNSVSGEVKTQRRKKRVAOPPVAYASAFPFLPAPSFYSP : 431									
ATPT8	-----LASOH----- : 387									

Figure 1

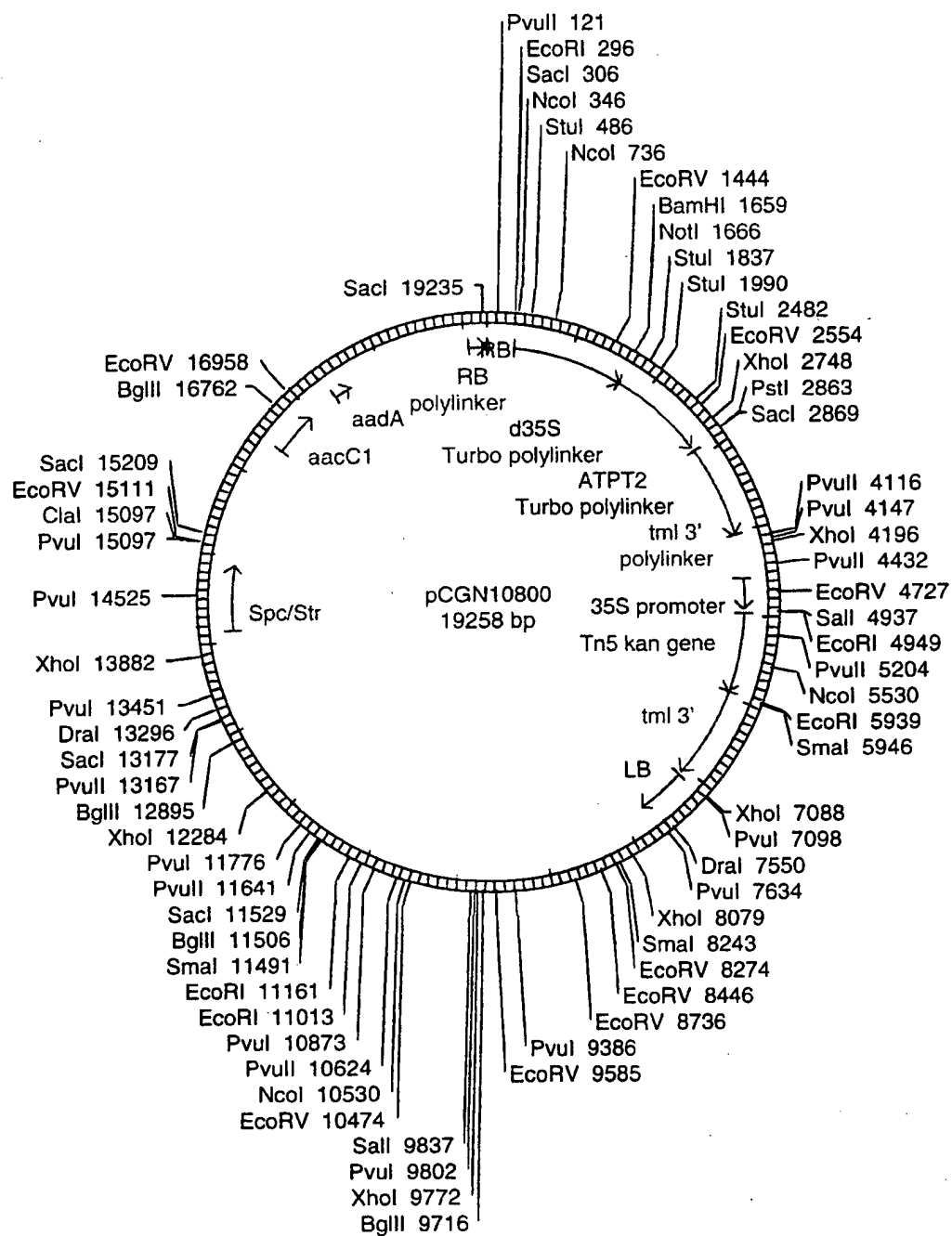


Figure 2

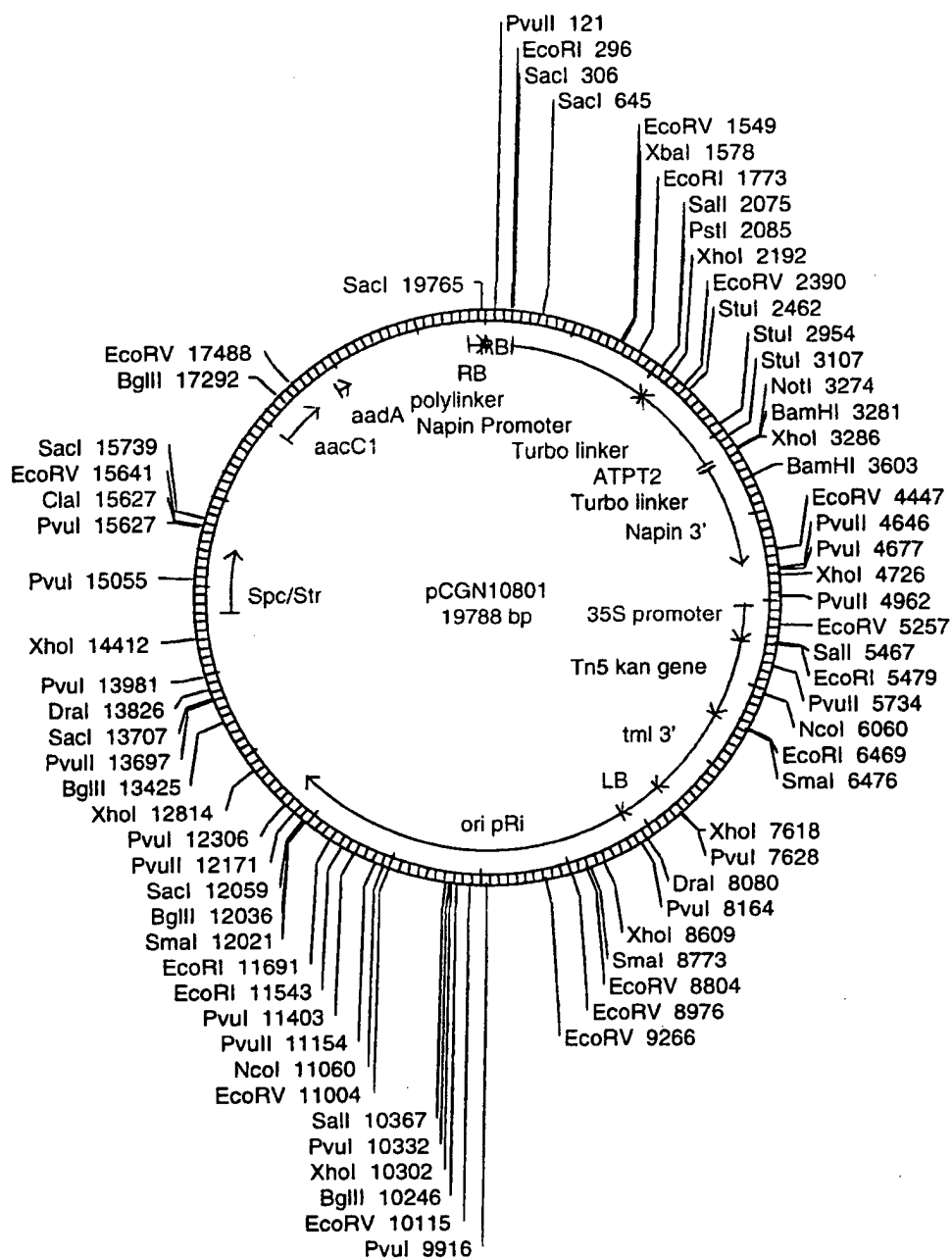


Figure 3

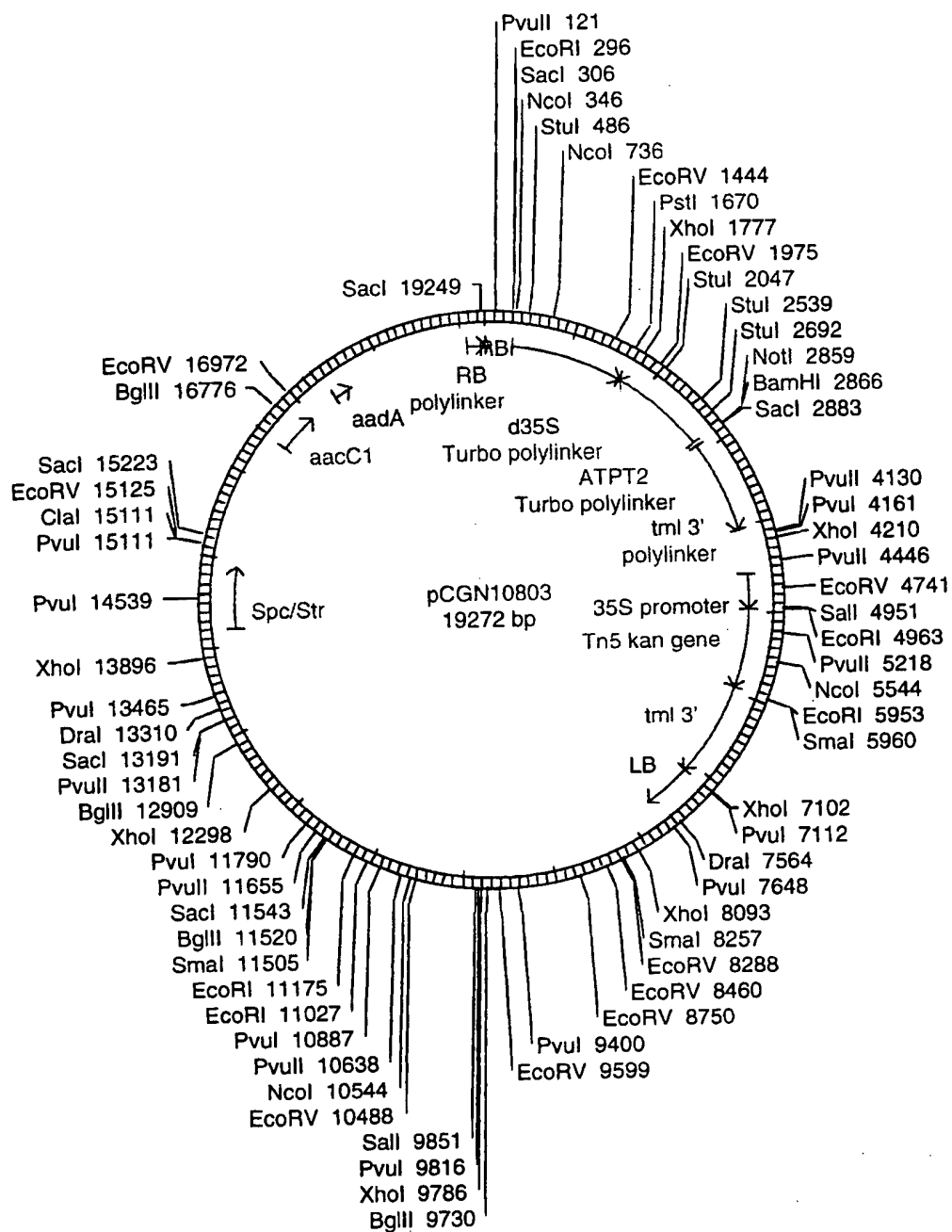


Figure 4

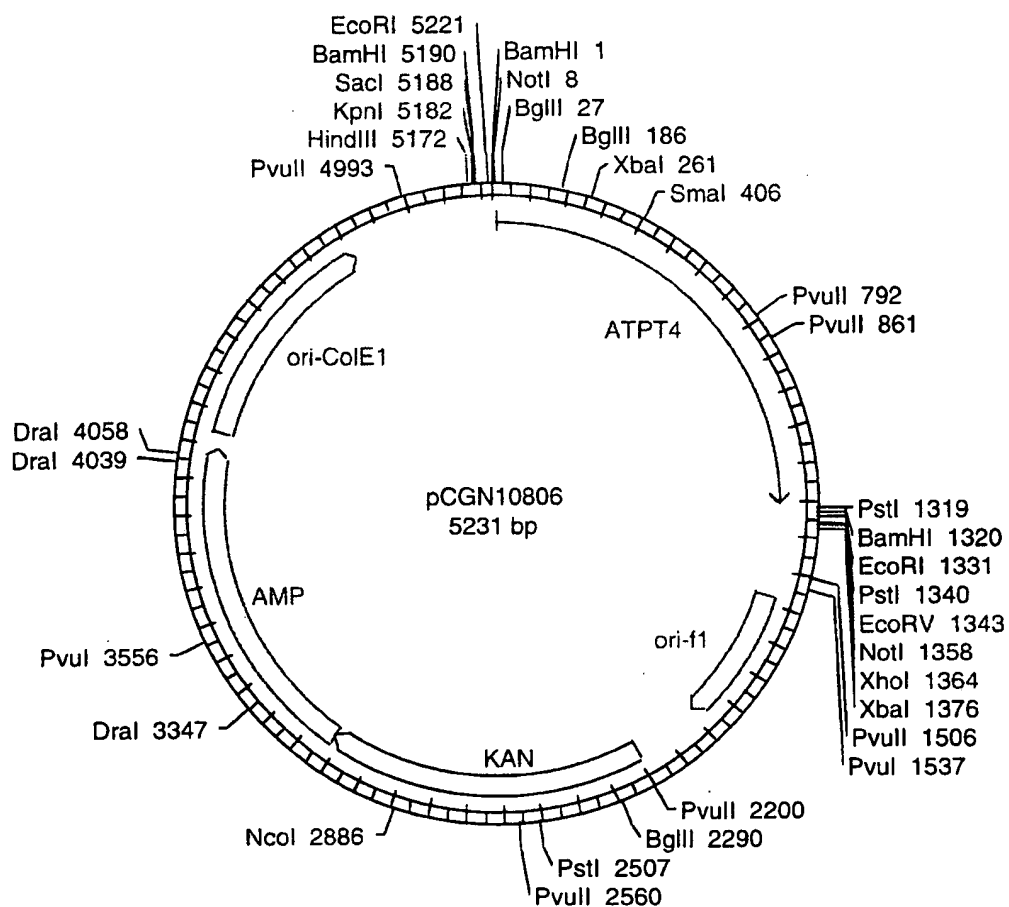


Figure 5

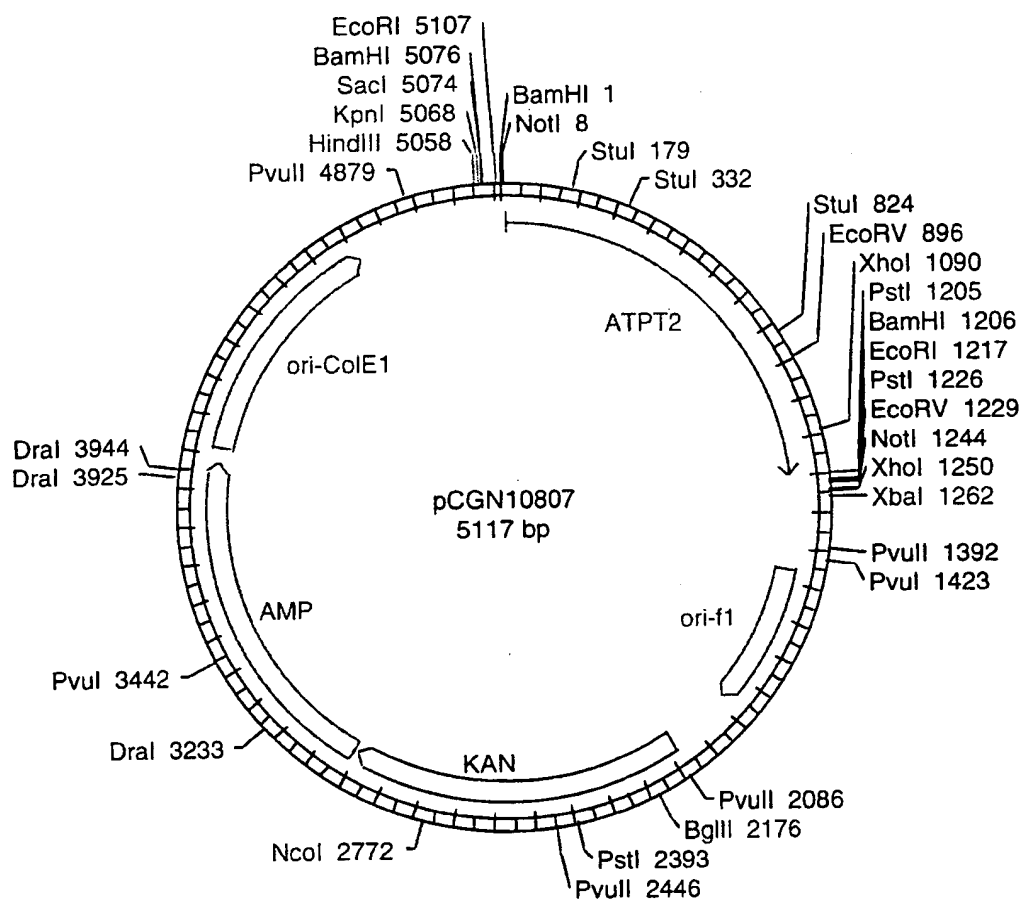


Figure 6

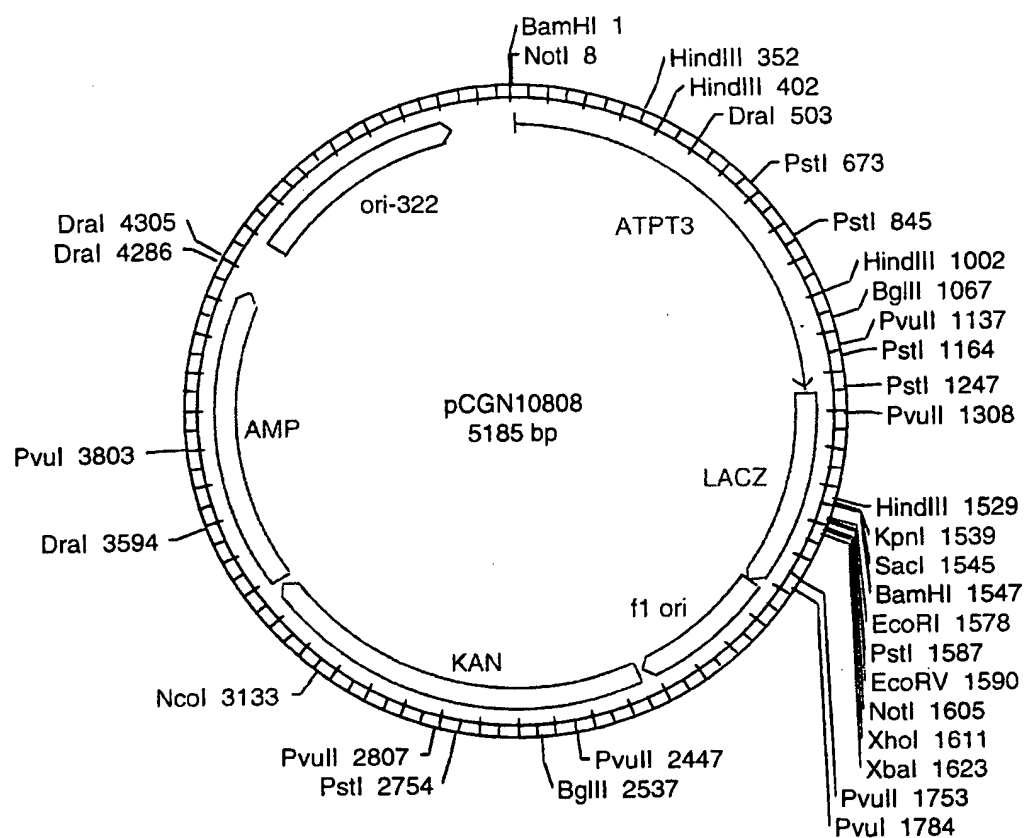


Figure 7

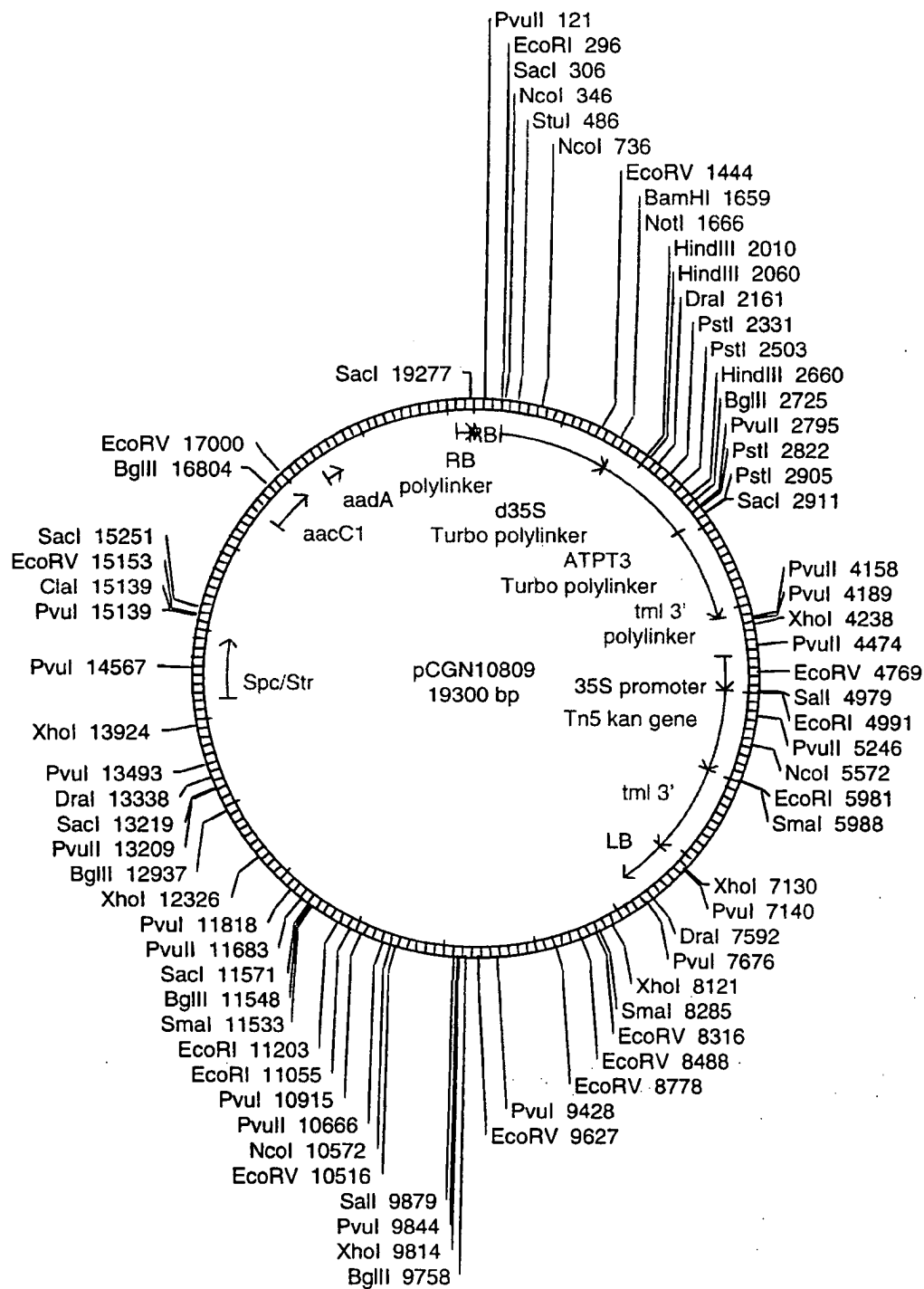


Figure 8

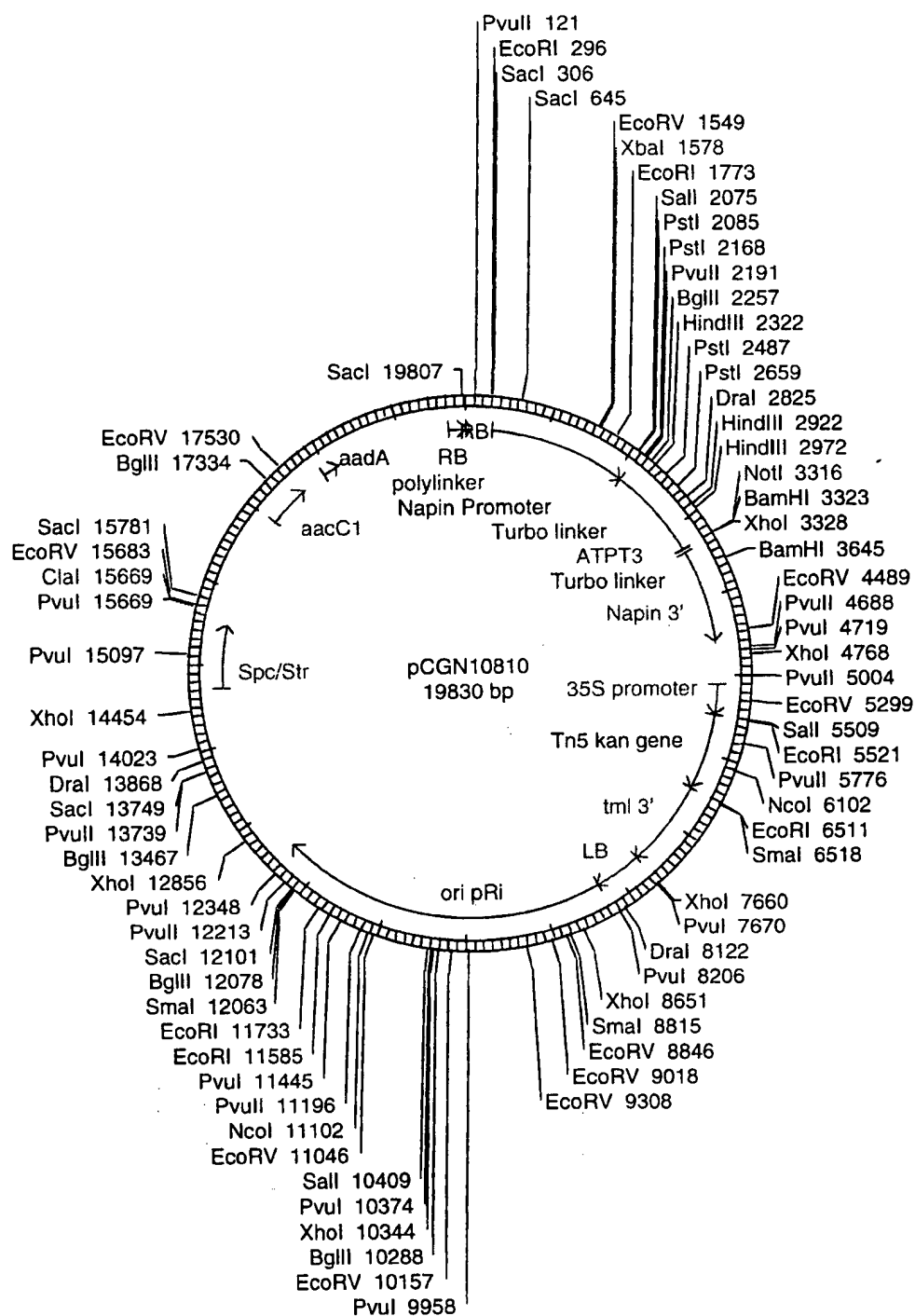


Figure 9

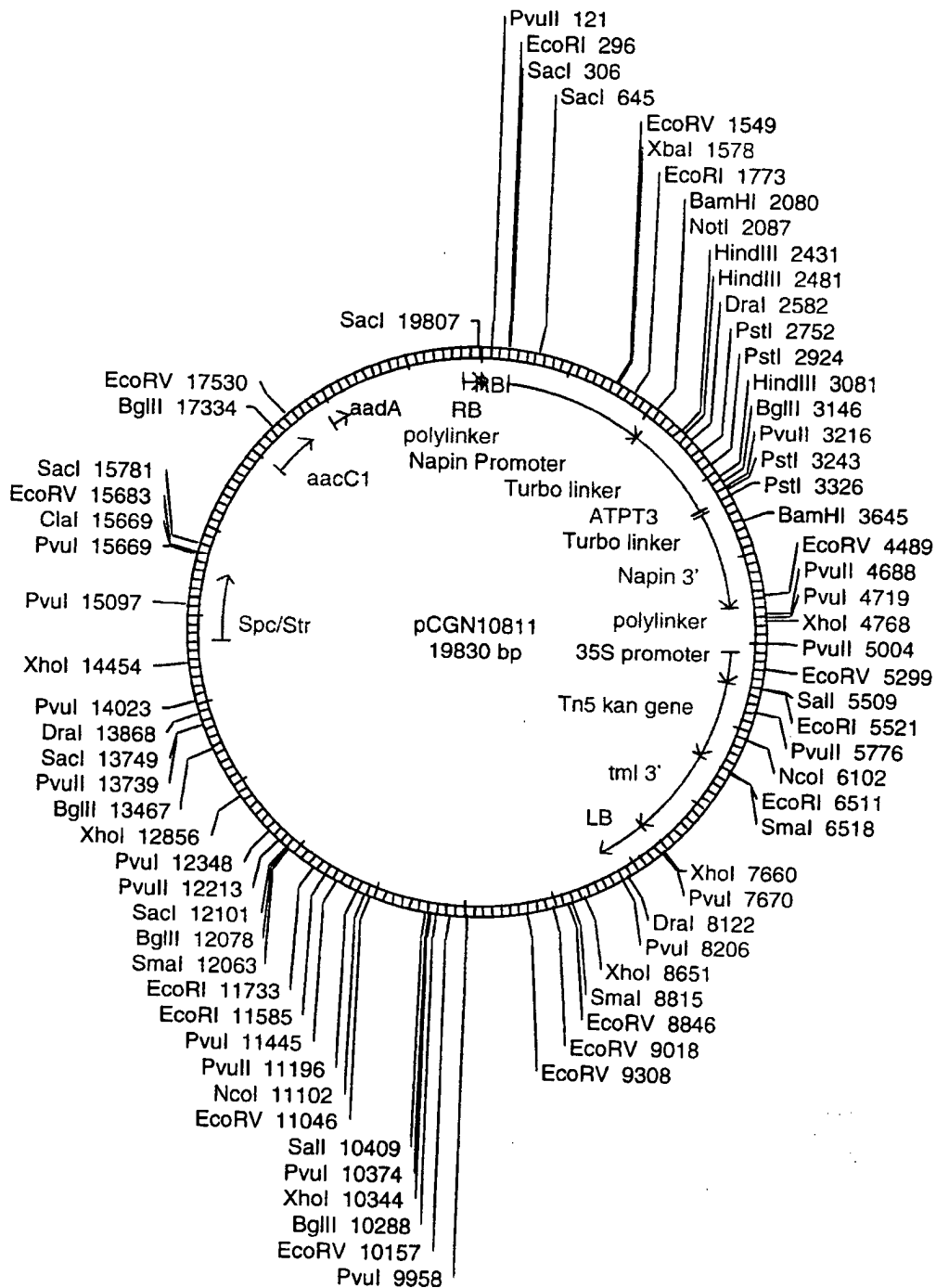


Figure 10

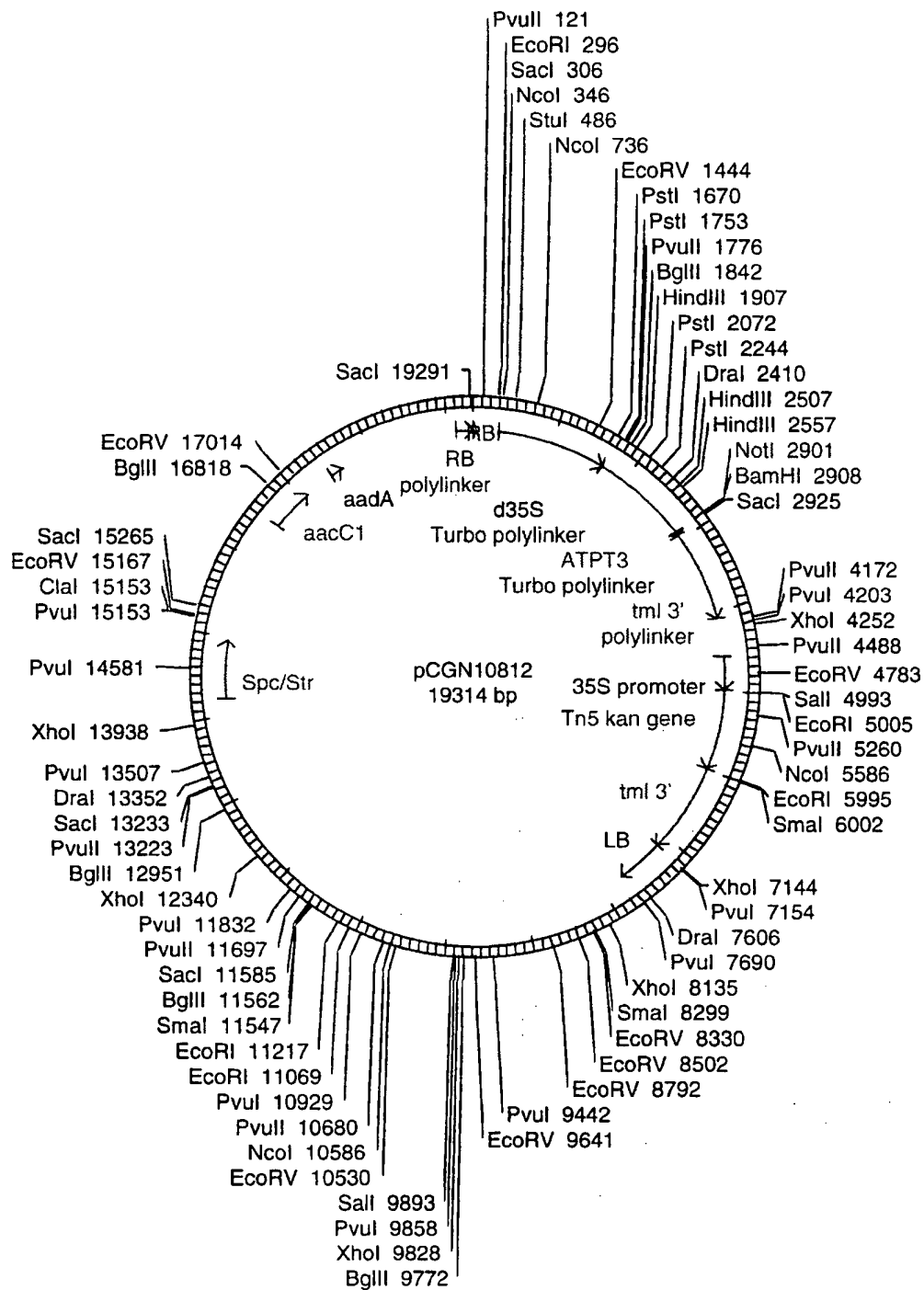


Figure 11

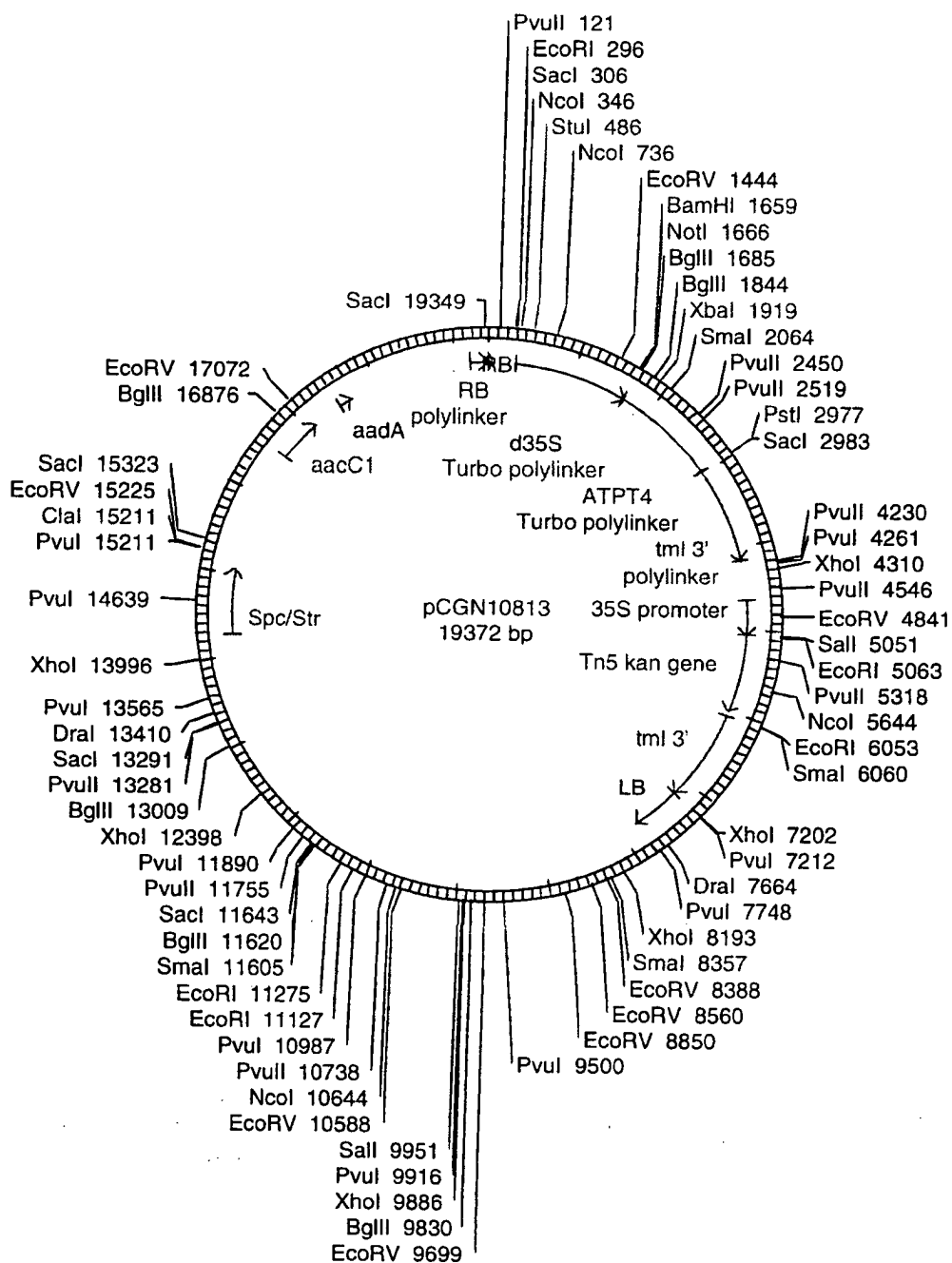


Figure 12

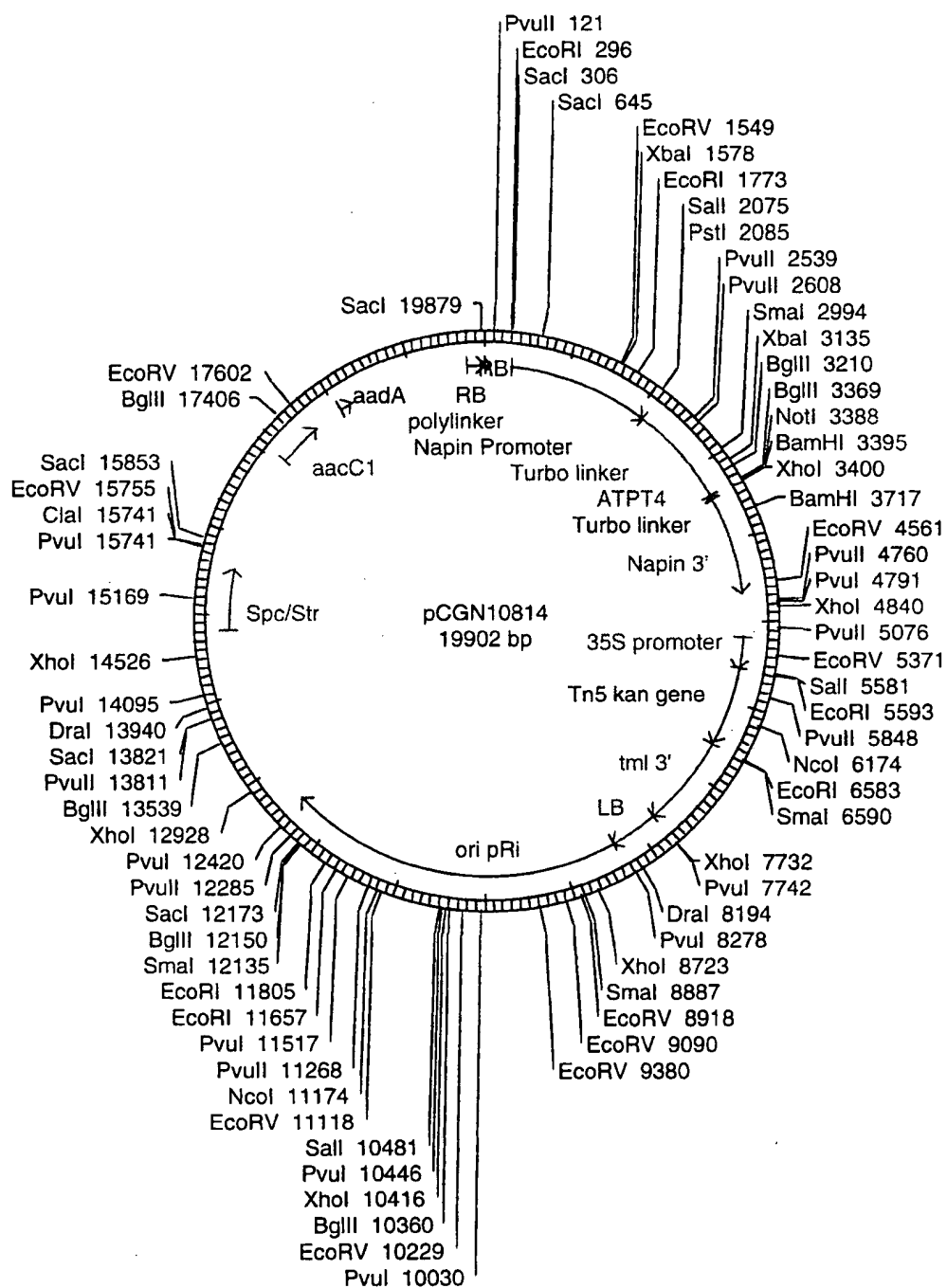


Figure 13

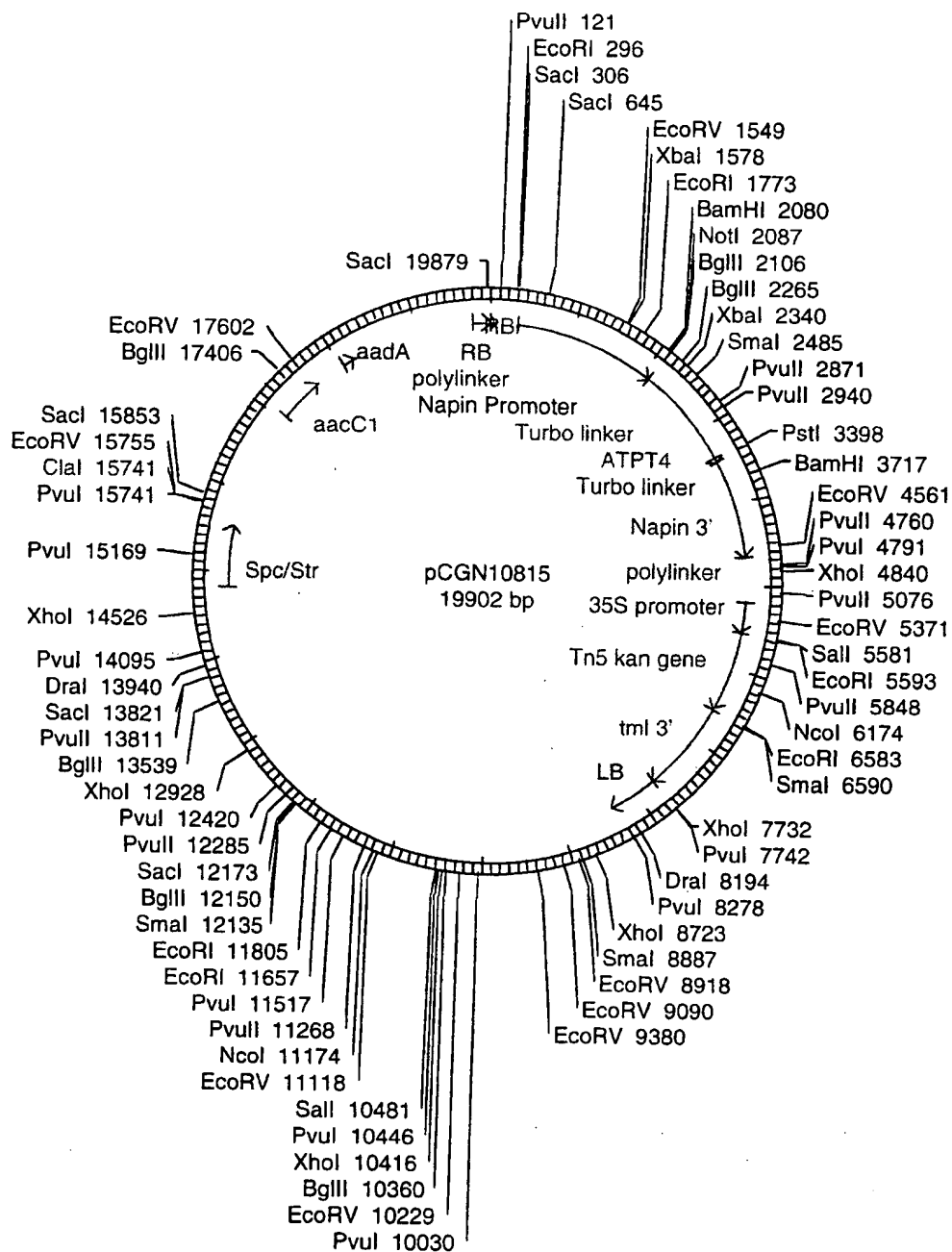


Figure 14

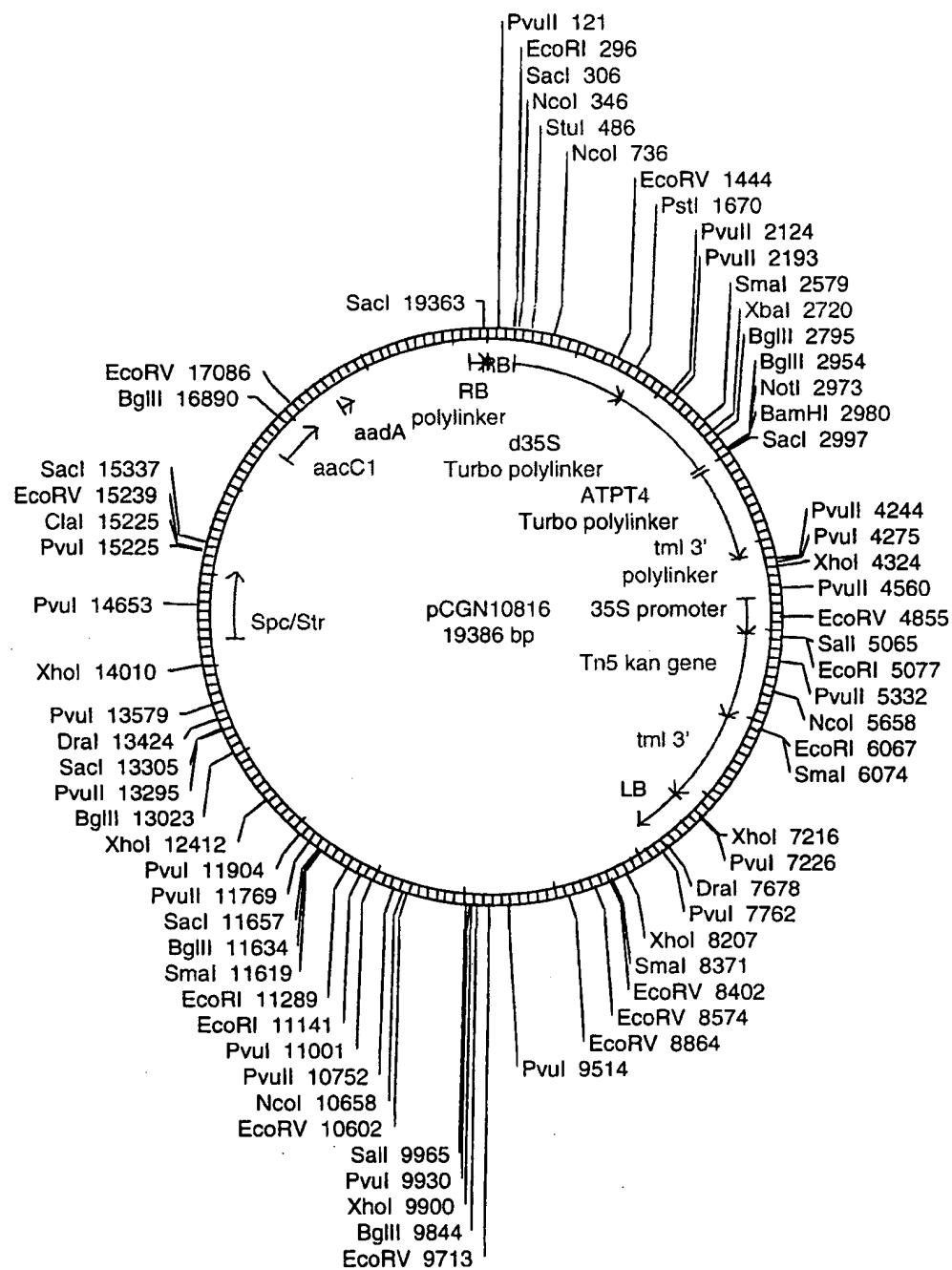


Figure 15

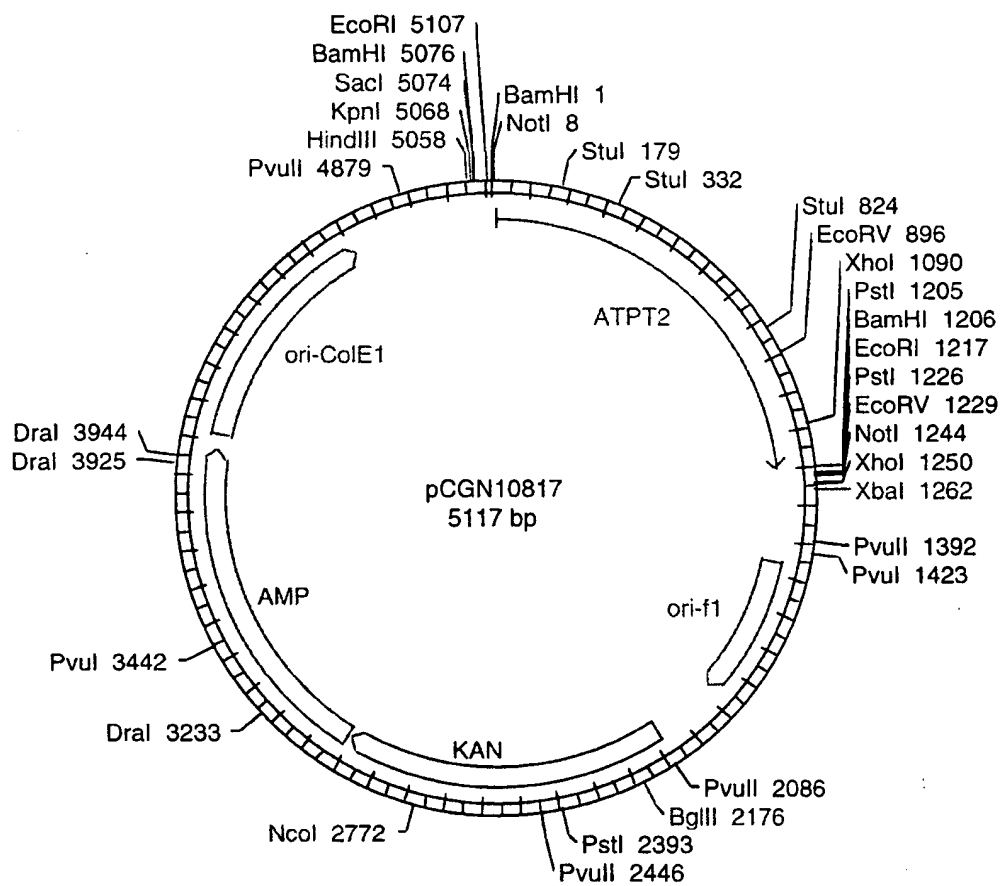


Figure 16

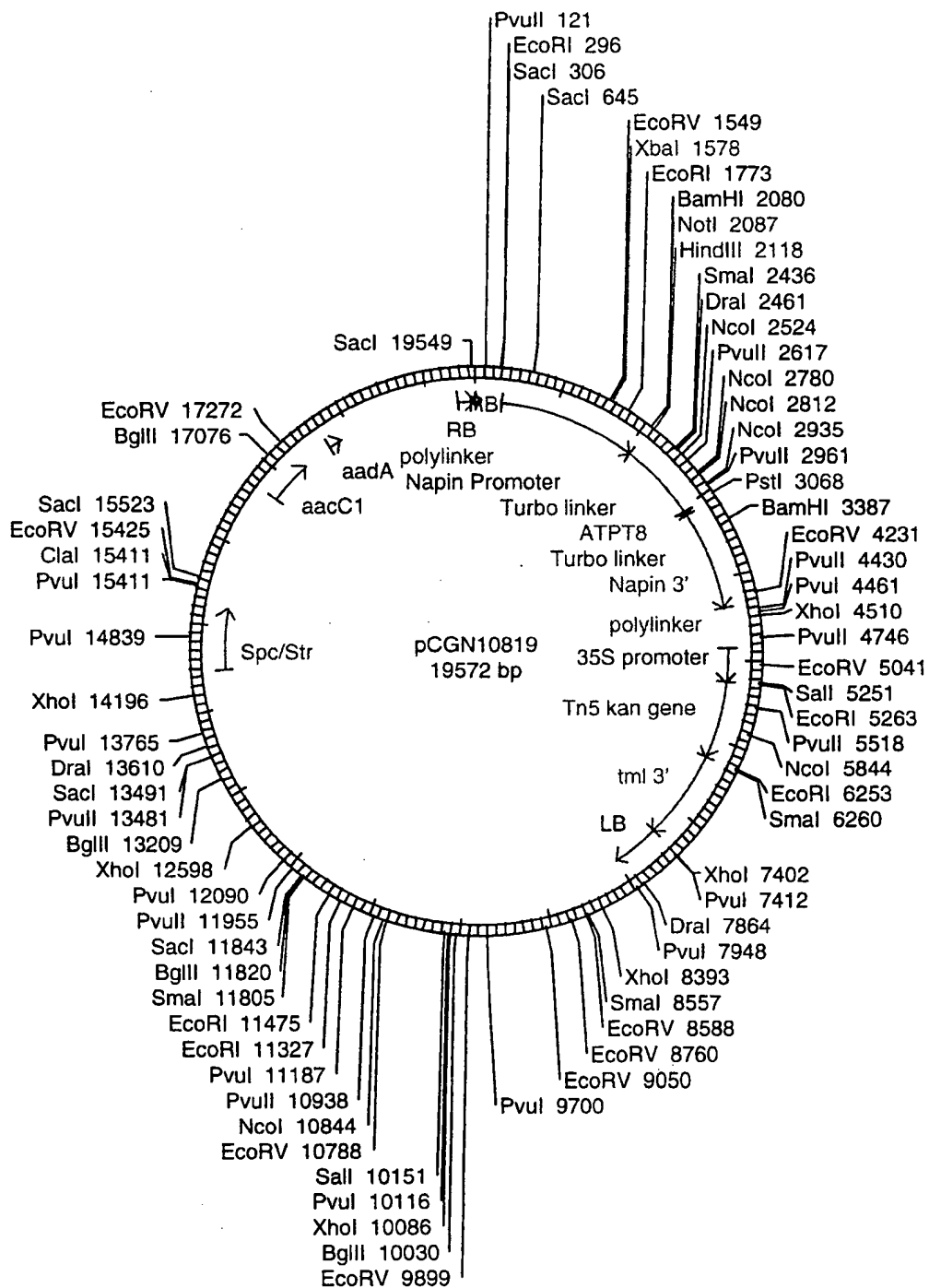


Figure 17

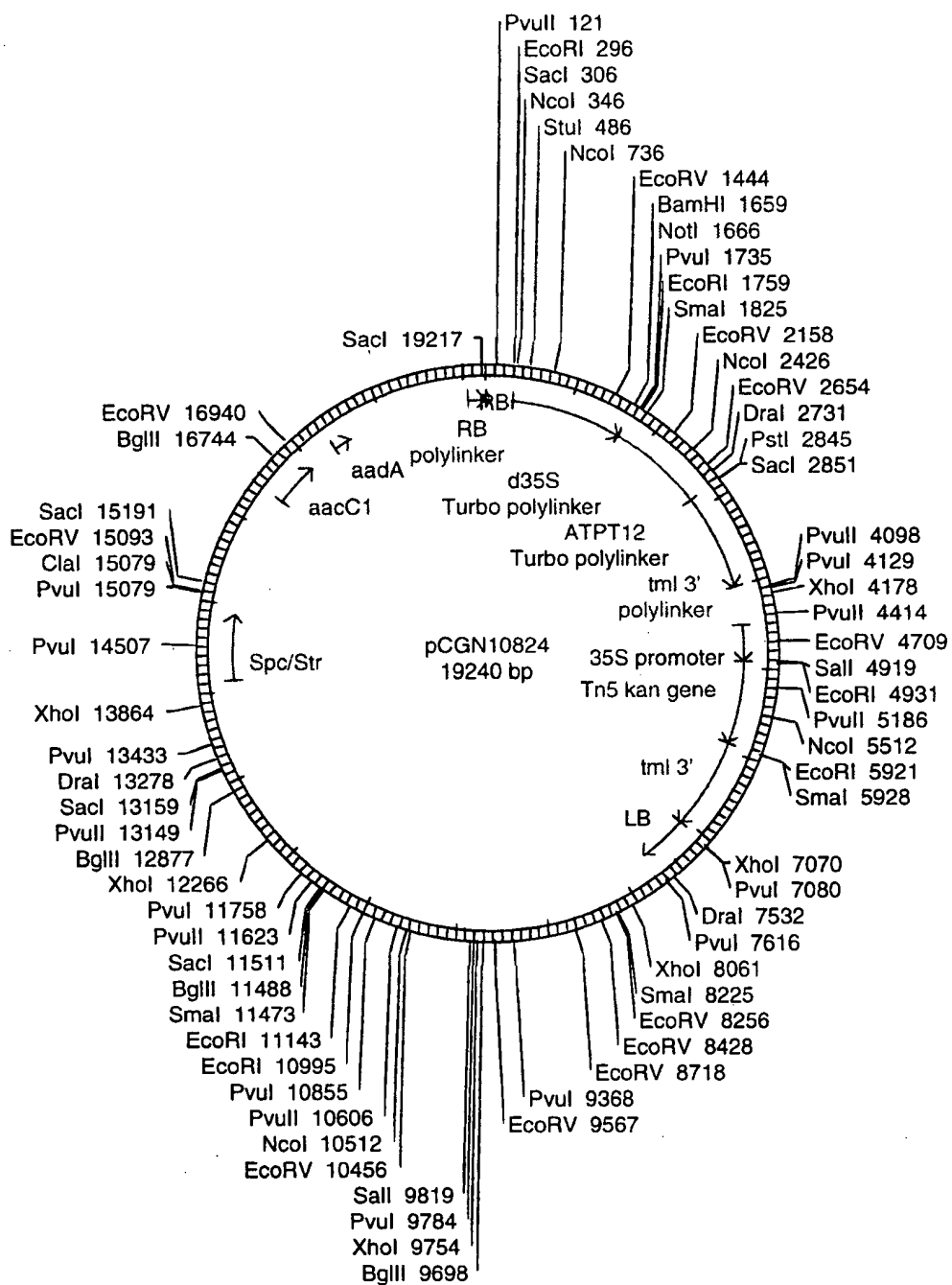


Figure 18

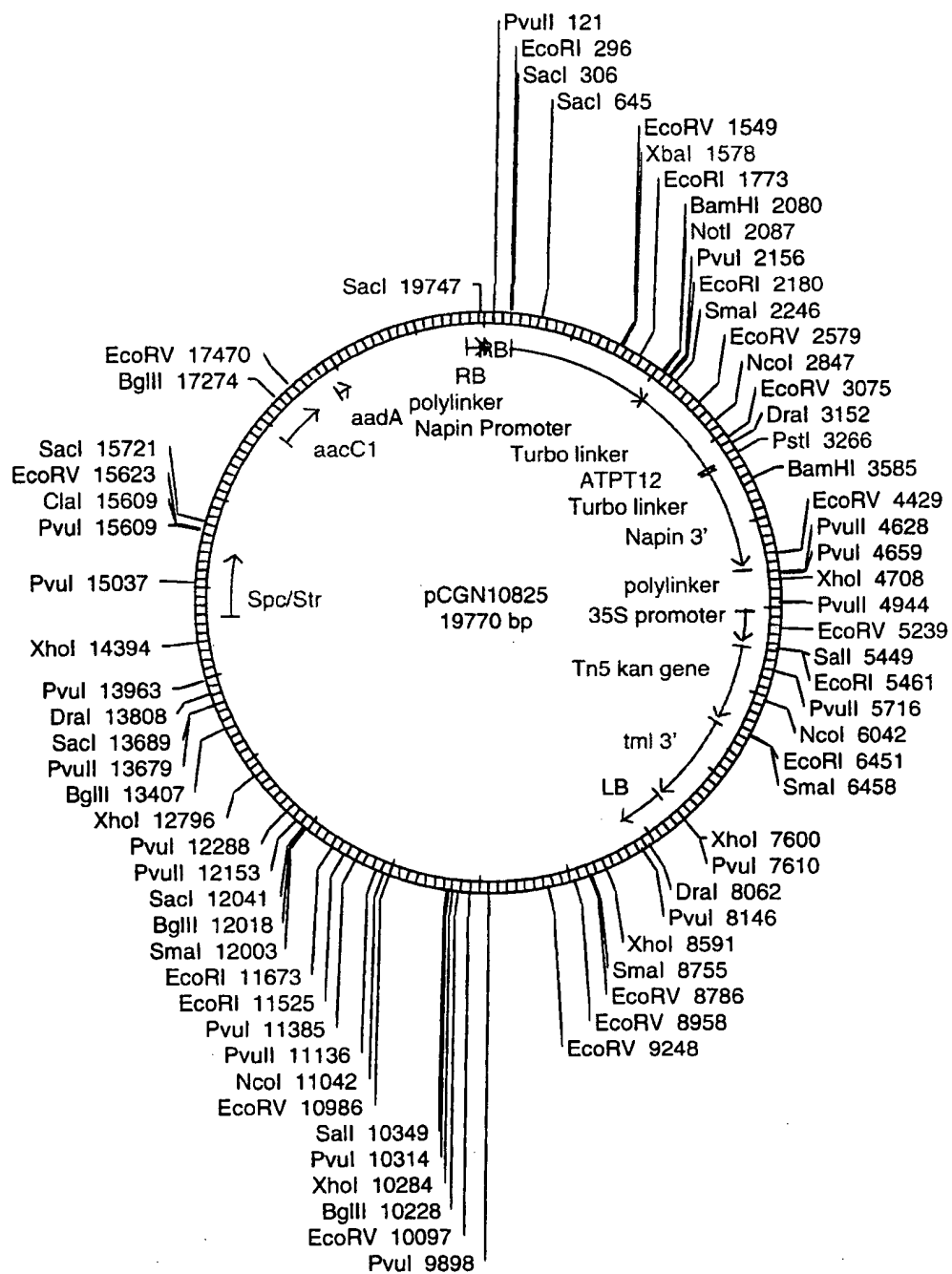


Figure 19

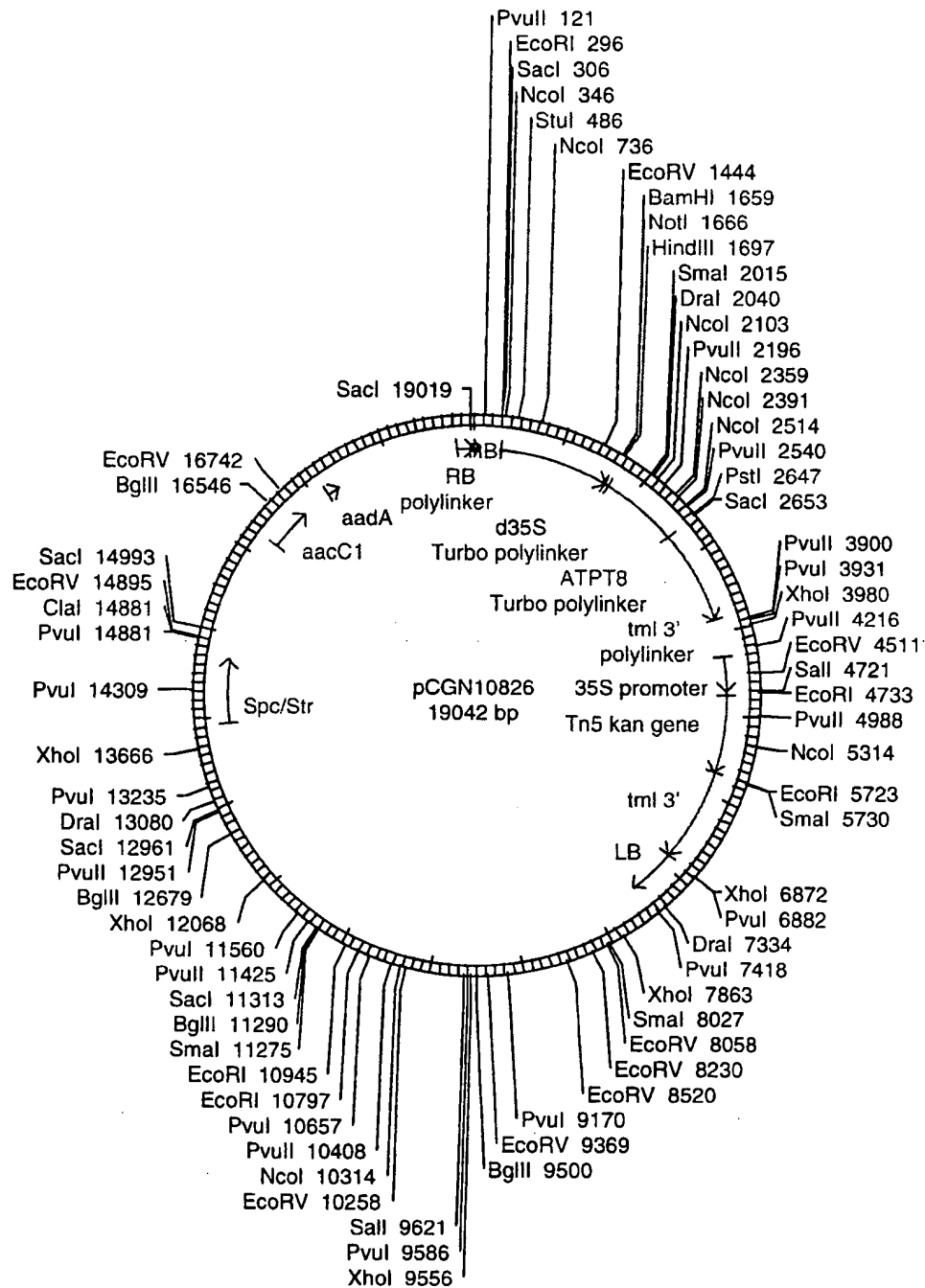


Figure 20

```

SLR1736 : MATIQAFWR : 20 * 40 * 60 80 * 152
SLR0926 : WVAQIPSS : PPLWLTIIYL : RWH : P : 64
SLR1899 : MVTSEKIHQHD : SMG : VCKSYQLT : P : 68
SLL1899 : MVTSEKIHQHD : SMG : VCKSYQLT : P : 71
SLR0056 : MSD : QNTCQ : NOAK : RQLIC : KGAAR : GESSIW : KRLQ : NKHT : TWIPL : WGV : CCAASS : GGYIWS : VEDFL : KAL : TCM : SGP : MTG : YQTL : 71
SLR1518 : MTESS : PLAP : STAP : ATR : KLM : AA : KP : P : 71

SLR1736 : NQEMVD : DR : INK : PNL : P : ANG : DFS : TAG : GRW : ING : CG : VAS : LA : IWG : GL : GL : TV : GIS : 152
SLR0926 : NDLMD : RDI : D : POW : ET : Q : R : LA : RA : L : S : OV : G : I : GA : VA : L : I : CA : A : CL : AF : T : PL : S : 150
SLL1899 : NCLVD : Q : D : I : YE : M : L : T : AR : I : P : AG : V : SPR : HA : LI : FA : L : V : SF : AL : AT : E : NV : LS : GC : 156
SLR0056 : NDFY : DR : D : I : DA : IN : E : PY : P : T : SG : I : SP : V : T : Q : L : L : VA : L : V : G : Y : C : DV : A : Q : HD : F : P : IM : MV : LT : C : A : F : AY : Y : S : AP : PL : K : L : Q : NG : W : I : GN : Y : A : L : 177
SLR1518 : NDFY : DS : D : I : G : I : D : V : RA : HS : V : N : LT : G : NR : NL : F : L : IS : NF : FL : A : G : V : I : GL : MS : S : WR : AQ : DM : 157

SLR1736 : VVNL : GL : F : T : F : R : I : GL : Y : PP : A : I : TP : A : V : I : H : F : L : M : E : T : VA : A : I : F : K : D : P : D : M : E : G : D : R : F : K : T : O : T : T : I : Q : I : G : N : Y : SV : PP : R : L : K : R : FS : L : LA : A : L : C : L : T : V : R : G : 241
SLR0926 : AWGE : AV : T : S : S : --- : A : V : G : D : I : D : A : T : W : I : M : G : A : T : V : F : W : I : G : F : D : T : V : Y : A : N : A : D : R : E : D : R : F : G : V : N : S : S : A : L : F : F : G : Y : G : E : A : V : G : F : F : A : 234
SLL1899 : AGS : I : PP : L : M : G : M : --- : A : V : D : I : S : W : TP : V : F : A : L : F : M : P : P : H : F : W : A : L : A : M : I : K : D : V : A : G : V : N : P : M : I : P : V : I : A : G : E : E : K : T : V : S : Q : I : W : Y : S : T : 241
SLR0056 : GAS : Y : I : A : P : W : I : G : H : A : U : F : G : I : N : P : T : M : V : L : A : I : Y : S : L : A : G : E : G : A : V : N : D : F : K : S : V : E : G : D : R : F : K : S : L : P : M : F : G : I : G : T : A : A : W : I : C : Y : M : 263
SLR1518 : LIT : F : GP : L : A : T : A : A : A : Y : S : Q : S : Q : S : F : S : W : N : L : TP : S : F : V : G : I : S : A : I : L : F : C : S : H : F : H : Q : V : E : D : L : A : A : K : K : S : P : I : V : R : L : G : T : K : G : S : Q : V : L : T : S : V : S : H : Y : I : T : A : I : G : V : I : C : H : 246

SLR1736 : AMPL : N : T : A : P : T : E : V : S : H : I : C : L : A : L : I : M : W : R : S : R : D : V : H : L : E : S : K : T : E : A : S : F : Y : Q : I : W : K : E : F : L : E : Y : L : Y : P : A : L : W : L : P : N : F : S : N : T : I : F : 308
SLR0926 : LNL : NP : V : M : S : A : T : A : G : M : V : I : Q : Y : I : L : S : A : P : T : E : P : K : L : Y : Q : - : I : G : Q : N : Y : I : G : F : V : I : L : A : G : M : L : G : W : I : 292
SLL1899 : LHQL : G : L : L : A : T : I : G : G : F : L : V : K : A : W : L : Q : K : A : P : G : D : R : D : I : A : R : G : L : K : F : S : I : Y : L : M : L : C : L : A : V : I : D : S : P : V : T : H : O : L : V : A : Q : M : G : T : L : L : G : 316
SLR0056 : YVH : Q : Q : Y : A : T : E : V : L : L : L : T : P : O : I : T : F : Q : D : M : Y : F : L : R : N : P : L : E : N : D : V : K : Y : Q : - : A : S : A : Q : P : F : L : V : F : G : M : L : A : T : G : A : L : G : H : A : G : I : 324
SLR1518 : QAP : W : Q : T : L : L : I : T : A : S : P : W : A : V : Q : L : I : R : H : V : G : Q : V : H : D : Q : E : Q : V : S : N : C : K : F : I : A : V : N : L : H : F : E : S : G : M : L : A : A : G : Y : W : A : G : I : G : 307

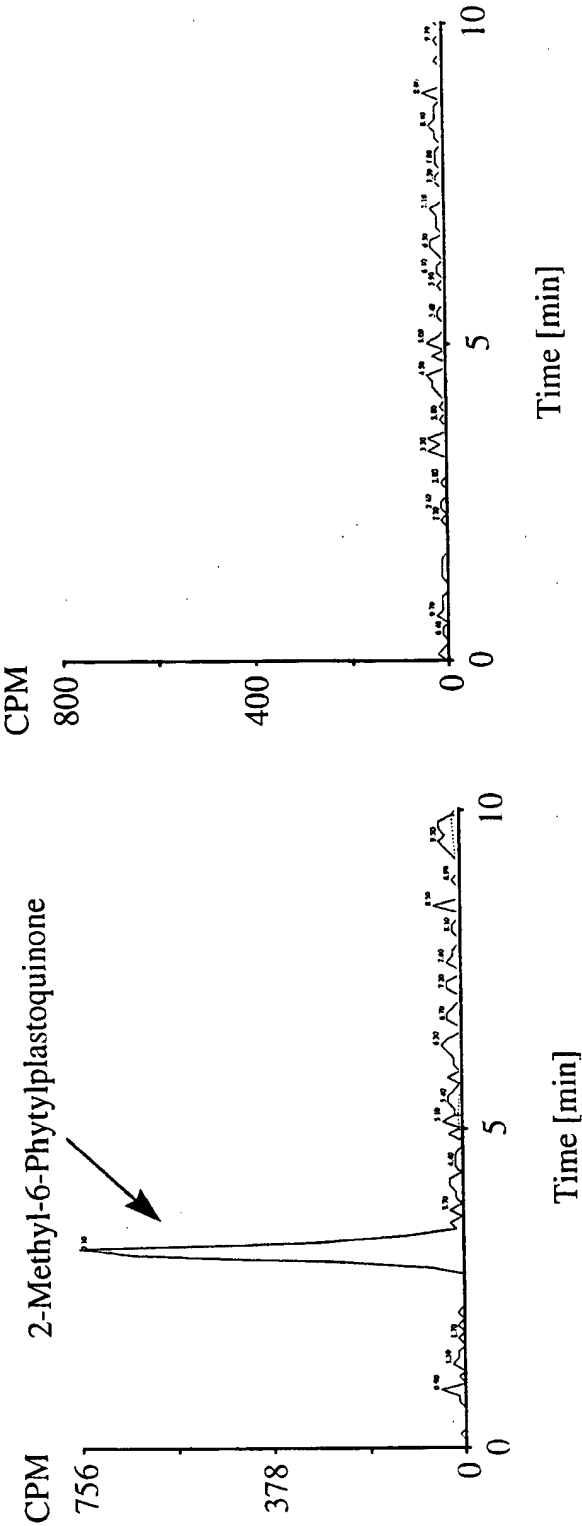
```

Figure 21

Figure 22 1/2

ATPT2	280	300	320	340	360	380	400	420	440	393
SLR1736	WKR-FALVAAMCILA	RAIIH-QIAFY	HH-IQTHVGRPI	IFTRPI	IFAFAMSF	SVVIA	FKDIPDI	GGKI-F	GRSES	TLG-QKR
ATPT3	LKR-FSLAALCILT	VRGIV-NLGLF	F-FRIGLGYPT	ITP	WV	TLFILV	VAIA	FKDVPDM	GRQ-FK	QNTLQIG-KQN
SLR0926	SIP-LMKRFTWPOA	FLGLT-NWGA	G-T	ATKSTAPSL	VP	YLSGVC	ILVYDTT	YAHQDK	DEVK	GKSTALRFG-DNT
ATPT4	AP-GAKRVPVPQ	LSIANGFAV	S--S--	ATSDATW	W	WGATVF	ALGFD	TVYAMADR	DDRR	GUNSSALRFG-QYV
SLR1899	VIT-PLKQLHPINT	WGAUGAIPP	G--A--	RASQI	YNSMI	PPAALYF	OLPHF	ALAHLCRNDY	AA-GYKMI	ISLFDP---S
ATPT12	VVTHLKRHTAQNIV	IGGAAGSIPP	AG--A--	ATGDL	WTPWV	FALIFL	PPHF	WALALMIKDDY	AAQ	NNPM
SLR0056	IS-APPLKQKQNGW	GNFAGASY	ISLPW	AGQ	ST	LPD	WV	TLIYSIAGI	GIA	VNDFKSV
ATPT8	IS-APPLKQKQNGW	GNFAGASY	ISLPW	AGQ	ST	LPD	WV	TLIYSIAGI	GIA	VNDFKSV
SLR1518	EITSSTEQRYSMDY	YKTYIKTAS	LSNSCKAV	ALITGO	AAE	AL	LAFEYGRNLG	LAFO	IDDI	LDFGTGSAS
	TKQPPFRLGLGELICL	ITFGP	AI-AAAYYSQSOSF	SWN	IT	PSVFVGIS	TAIILFC	SHFHQV	DDLA-A	KKSP
ATPT2	360	380	400	420	440	460	480	500	520	537
SLR1736	VFWTC	TLQ	MAVAY	AI	VGATSP	FIWSKV	ISV	GHVIT	ATTT	WARAKSV
ATPT3	KLW	TGFTASIG	FALS	GF	SADLG	WQYYAS	AAASQ	GWQI	GTADL	SSGADCS
SLR0926	GEAV	GFFALTIC	FY	GM	MLNPL	YWLST	ATAI	GGW	IQI	LSAPTPEP-KY
ATPT4	GKR	AAVA	RNC	FY	IP	GF	AYDWGL	TSSWFC	ESTL	TLA
SLR1899	VSO	WYYS	LVP	PSLL	VY	PLHQL	GILYLA	AAI	GGQ	FLVKA
ATPT12	AKW	ICGA	DIT	Q	SVAGYL	ASGKPY	YALAL	VAL	PPQ	VFQKY
SLR0056	AAW	ICIM	DVF	QAGI	AGYL	VVHQ	Q	LYATIV	LL	PPQ
ATPT8	ITAP	ILFAMEEF	PQ	RE	VDO	VEKDP	RNV	DIA	LEY	GKSKG
SLR1518	GSQ	VL	LSV	LSLYL	TA	GV	CHQAP	WQTL	LL	IAS
ATPT2	NTIF	---	---	---	---	---	---	---	---	---
SLR1736	---	---	---	---	---	---	---	---	---	---
ATPT3	---	---	---	---	---	---	---	---	---	---
SLR0926	---	---	---	---	---	---	---	---	---	---
ATPT4	---	---	---	---	---	---	---	---	---	---
SLR1899	---	---	---	---	---	---	---	---	---	---
ATPT12	---	---	---	---	---	---	---	---	---	---
SLR0056	---	---	---	---	---	---	---	---	---	---
ATPT8	---	---	---	---	---	---	---	---	---	---
SLR1518	---	---	---	---	---	---	---	---	---	---

Figure 22 2/2



Synechocystis 6803 wild type *Synechocystis* slr1736 knockout

Figure 23

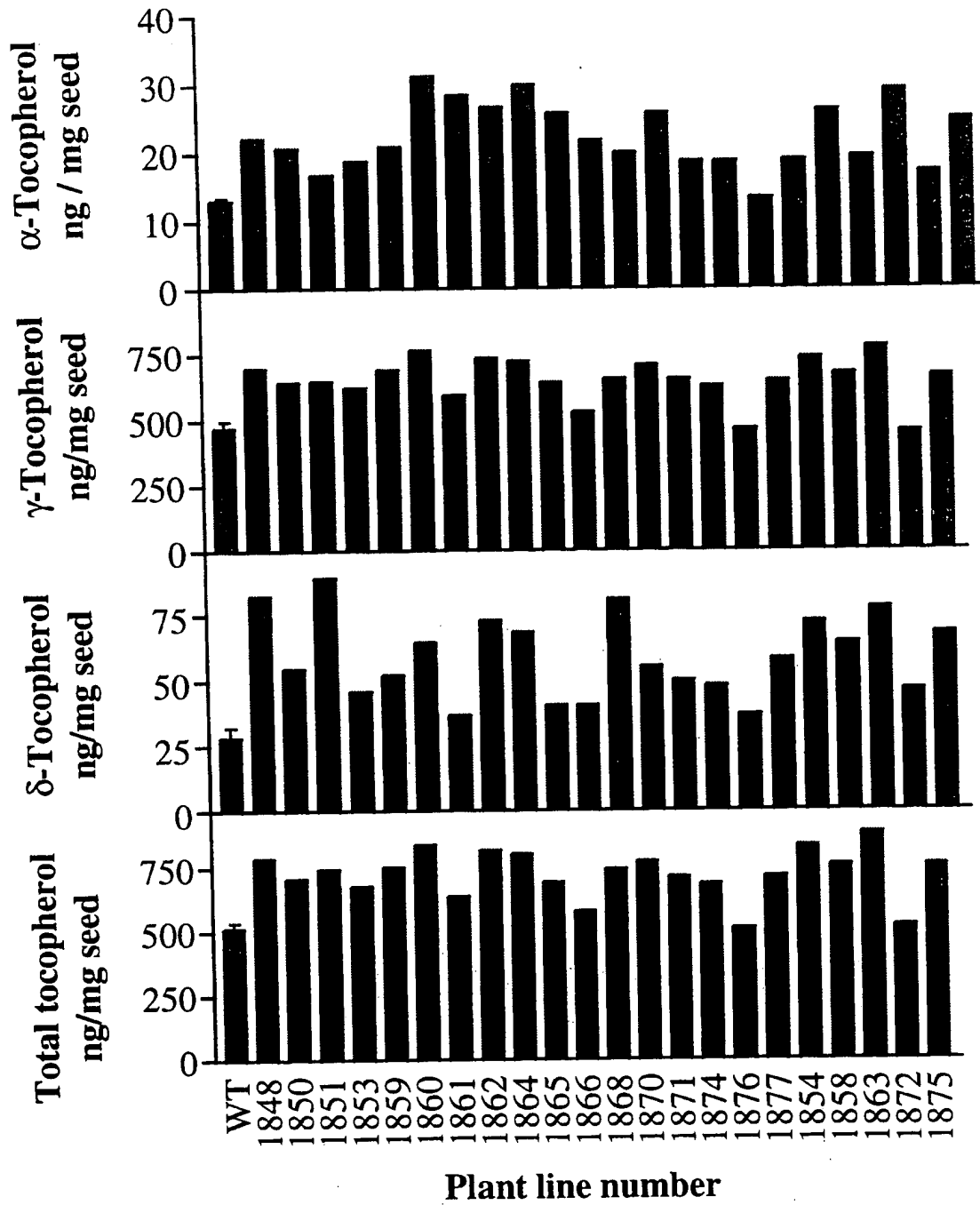


Figure 24

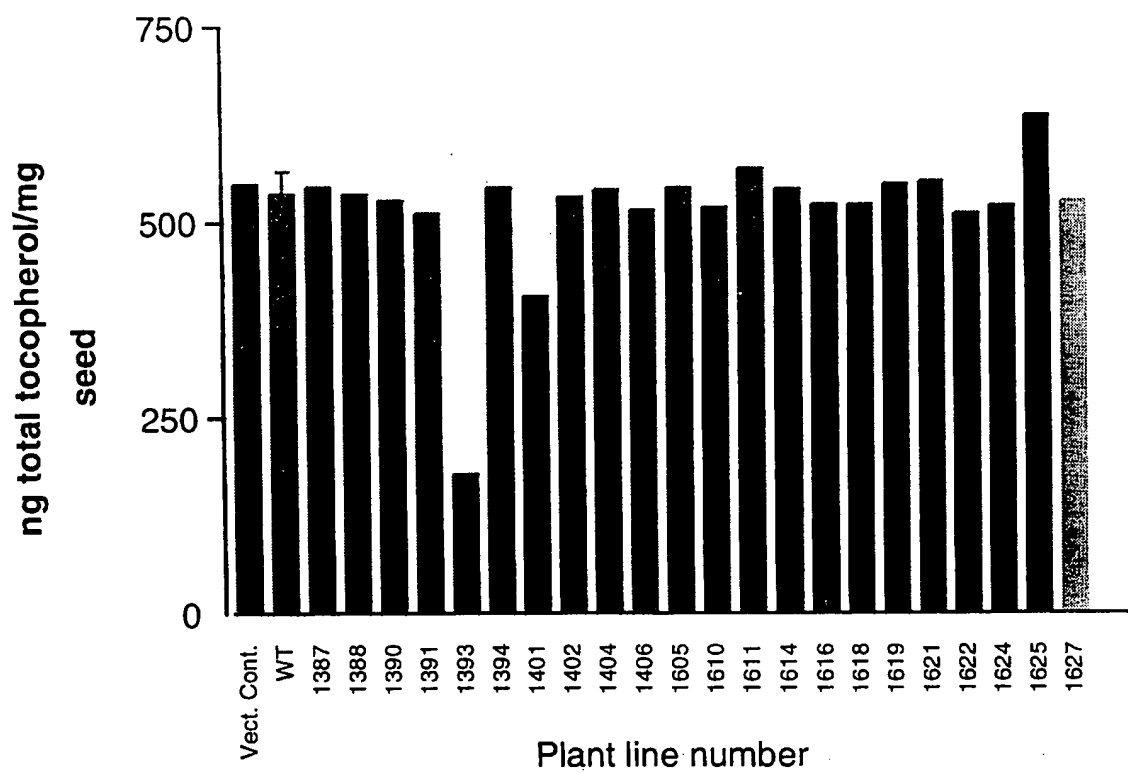


Figure 25

SEQUENCE LISTING

5 <110> Calgene LLC
 <120> Nucleic Acid Sequences Involved in
 Tocopherol Synthesis
 <130> 17133/00/WO
 10 <150> 60/129,899
 <151> 1999-04-15
 <150> 60/146,461
 15 <151> 1999-07-30
 <160> 94
 <170> FastSEQ for Windows Version 4.0
 20 <210> 1
 <211> 1182
 <212> DNA
 <213> Arabidopsis sp
 25 <400> 1
 atggagtcctc tgctctctag ttcttctctt gtttccgctg ctggtgggtt ttgttggaag 60
 aagcagaatc taaagctcca ctctttatca gaaatccgag ttctgcgttg tgattcgagt 120
 aaagtgtgctg caaaaccgaa gtttaggaac aatcttggtta ggctgatgg tcaaggatct 180
 30 tcattgttgt tgtatccaaa acataagtcg agatttcggg ttaatgccac tgcgggtcag 240
 cctgaggctt tcgactcgaa tagcaaacag aagtctttta gagactcgtt agatgcgttt 300
 tacaggtttt ctaggcctca tacagttatt ggcacagtgc ttagcatttt atctgtatct 360
 ttcttagcag tagagaaggt ttctgatata tctcctttac ttttactgg catcttggag 420
 gctgttgttg cagctctcat gatgaacatt tacatagttg ggctaaatca gttgtctgat 480
 35 gttgaaatag ataagggttaa caagccctat cttccattgg catcaggaga atattctggt 540
 aacaccggca ttgcaatagt agcttccttc tccatcatga gtttctggct tgggtggatt 600
 gttgggttcat ggccattggt ctgggctctt tttgtgagtt tcatgctcgg tactgcatac 660
 tctatcaatt tgccactttt acggtggaaa agatttgcag tggttgcagc aatgtgtatc 720
 ctgcgtgtcc gagctattat tgttcaaata gcctttttatc tacatattca gacacatgtg 780
 40 tttggaagac caatcttggt cactaggcct cttattttcg ccactgcgtt tatgagcttt 840
 ttctctgtcg ttattgcatt gtttaaggat atacctgata tcgaagggga taagatatc 900
 ggaatccgat cattctctgt aactctgggt cagaaacggg tgttttggac atgtgttaca 960
 ctacttcaaa tggcttacgc tgttgcaatt ctagtggag ccacatctcc attcatatgg 1020
 agcaaagtca tctcggttgt gggctcatgt atactcgcaa caactttgtg ggctcgagct 1080
 45 aagtccgttg atctgagtag caaaaccgaa ataacttcat gttatatgtt catatggaag 1140
 ctctttttatg cagagtactt gctgttacct tttttgaagt ga 1182

	<400> 2															
	Met	Glu	Ser	Leu	Leu	Ser	Ser	Ser	Ser	Leu	Val	Ser	Ala	Ala	Gly	Gly
	1				5					10					15	
10	Phe	Cys	Trp	Lys	Lys	Gln	Asn	Leu	Lys	Leu	His	Ser	Leu	Ser	Glu	Ile
				20					25					30		
	Arg	Val	Leu	Arg	Cys	Asp	Ser	Ser	Lys	Val	Val	Ala	Lys	Pro	Lys	Phe
				35					40					45		
	Arg	Asn	Asn	Leu	Val	Arg	Pro	Asp	Gly	Gln	Gly	Ser	Ser	Leu	Leu	Leu
15		50					55					60				
	Tyr	Pro	Lys	His	Lys	Ser	Arg	Phe	Arg	Val	Asn	Ala	Thr	Ala	Gly	Gln
	65					70					75					80
	Pro	Glu	Ala	Phe	Asp	Ser	Asn	Ser	Lys	Gln	Lys	Ser	Phe	Arg	Asp	Ser
					85					90					95	
20	Leu	Asp	Ala	Phe	Tyr	Arg	Phe	Ser	Arg	Pro	His	Thr	Val	Ile	Gly	Thr
				100					105					110		
	Val	Leu	Ser	Ile	Leu	Ser	Val	Ser	Phe	Leu	Ala	Val	Glu	Lys	Val	Ser
				115					120					125		
	Asp	Ile	Ser	Pro	Leu	Leu	Phe	Thr	Gly	Ile	Leu	Glu	Ala	Val	Val	Ala
25		130					135					140				
	Ala	Leu	Met	Met	Asn	Ile	Tyr	Ile	Val	Gly	Leu	Asn	Gln	Leu	Ser	Asp
	145					150					155					160
	Val	Glu	Ile	Asp	Lys	Val	Asn	Lys	Pro	Tyr	Leu	Pro	Leu	Ala	Ser	Gly
					165					170					175	
30	Glu	Tyr	Ser	Val	Asn	Thr	Gly	Ile	Ala	Ile	Val	Ala	Ser	Phe	Ser	Ile
				180					185					190		
	Met	Ser	Phe	Trp	Leu	Gly	Trp	Ile	Val	Gly	Ser	Trp	Pro	Leu	Phe	Trp
				195					200					205		
	Ala	Leu	Phe	Val	Ser	Phe	Met	Leu	Gly	Thr	Ala	Tyr	Ser	Ile	Asn	Leu
35		210					215					220				
	Pro	Leu	Leu	Arg	Trp	Lys	Arg	Phe	Ala	Leu	Val	Ala	Ala	Met	Cys	Ile
	225					230					235					240
	Leu	Ala	Val	Arg	Ala	Ile	Ile	Val	Gln	Ile	Ala	Phe	Tyr	Leu	His	Ile
					245					250					255	
40	Gln	Thr	His	Val	Phe	Gly	Arg	Pro	Ile	Leu	Phe	Thr	Arg	Pro	Leu	Ile
				260					265					270		
	Phe	Ala	Thr	Ala	Phe	Met	Ser	Phe	Phe	Ser	Val	Val	Ile	Ala	Leu	Phe
				275				280					285			
	Lys	Asp	Ile	Pro	Asp	Ile	Glu	Gly	Asp	Lys	Ile	Phe	Gly	Ile	Arg	Ser
45		290					295					300				
	Phe	Ser	Val	Thr	Leu	Gly	Gln	Lys	Arg	Val	Phe	Trp	Thr	Cys	Val	Thr

305 310 315 320
 Leu Leu Gln Met Ala Tyr Ala Val Ala Ile Leu Val Gly Ala Thr Ser
 325 330 335
 Pro Phe Ile Trp Ser Lys Val Ile Ser Val Val Gly His Val Ile Leu
 5 340 345 350
 Ala Thr Thr Leu Trp Ala Arg Ala Lys Ser Val Asp Leu Ser Ser Lys
 355 360 365
 Thr Glu Ile Thr Ser Cys Tyr Met Phe Ile Trp Lys Leu Phe Tyr Ala
 370 375 380
 10 Glu Tyr Leu Leu Leu Pro Phe Leu Lys
 385 390

 <210> 3
 <211> 1224
 15 <212> DNA
 <213> Arabidopsis sp

 <400> 3
 20 atggcggtttt ttgggctctc ccgtgttttca agacggttgt tgaaatcttc cgtctccgta 60
 actccatctt ctctctctgc tcttttgcaa tcacaacata aatccttgtc caatcctgtg 120
 actaccattt acacaaatcc tttcactaag tgttatcctt catggaatga taattaccaa 180
 gtatggagta aaggaagaga attgcatcag gagaagtttt ttgggtgttg ttggaattac 240
 agattaattt gtggaatgtc gtcgtcttct tcggttttgg agggaaagcc gaagaaagat 300
 gataaggaga agagtgatgg tgttgttgtt aagaaagctt cttggataga tttgtattta 360
 25 ccagaagaag ttagagggtta tgctaagctt gtcgatttgg ataaacccat tggaacttgg 420
 ttgcttgcgt ggccttgtat gtggtcgatt gcgttggctg ctgatcctgg aagccttcca 480
 agttttaaat atatggcttt atttggttgc ggagcattac ttcttagagg tgctgggtgt 540
 actataaatg atctgcttga tcaggacata gatacaaagg ttgatcgtac aaaactaaga 600
 cctatcgcca gtggtctttt gacaccattt caagggtattg gatttctcgg gctgcagttg 660
 30 ctttttaggt tagggattct tctccaactt aacaattaca gccgtgtttt aggggcttca 720
 tctttgttac ttgtcttttc ctaccactt atgaagaggt ttacattttg gcctcaagcc 780
 ttttttaggt tgaccataaa ctggggagca ttgttaggat ggactgcagt taaaggaagc 840
 atagcaccat ctattgtact ccctctctat ctctccggag tctgctggac ccttgtttat 900
 gatactattt atgcacatca ggacaaagaa gatgatgtaa aagttgggtg taagtcaaca 960
 35 gcccttagat tcggtgataa tacaaagctt tgggttaactg gatttggcac agcatccata 1020
 ggttttcttg cactttctgg attcagtga gatctcgggt ggcaatatta cgcactactg 1080
 gccgctgcat caggacagtt aggatggcaa atagggacag ctgacttata atctggtgct 1140
 gactgcagta gaaaatttgt gtcgaacaag tgggttgggt ctattatatt tagtggagtt 1200
 gtacttgga gaagttttca ataa 1224
 40

 <210> 4
 <211> 407
 <212> PRT
 <213> Arabidopsis sp
 45

 <400> 4

	Met	Ala	Phe	Phe	Gly	Leu	Ser	Arg	Val	Ser	Arg	Arg	Leu	Leu	Lys	Ser
	1				5				10						15	
	Ser	Val	Ser	Val	Thr	Pro	Ser	Ser	Ser	Ser	Ala	Leu	Leu	Gln	Ser	Gln
				20				25					30			
5	His	Lys	Ser	Leu	Ser	Asn	Pro	Val	Thr	Thr	His	Tyr	Thr	Asn	Pro	Phe
			35				40					45				
	Thr	Lys	Cys	Tyr	Pro	Ser	Trp	Asn	Asp	Asn	Tyr	Gln	Val	Trp	Ser	Lys
		50					55				60					
	Gly	Arg	Glu	Leu	His	Gln	Glu	Lys	Phe	Phe	Gly	Val	Gly	Trp	Asn	Tyr
10	65					70					75				80	
	Arg	Leu	Ile	Cys	Gly	Met	Ser	Ser	Ser	Ser	Ser	Val	Leu	Glu	Gly	Lys
				85					90						95	
	Pro	Lys	Lys	Asp	Asp	Lys	Glu	Lys	Ser	Asp	Gly	Val	Val	Val	Lys	Lys
				100					105						110	
15	Ala	Ser	Trp	Ile	Asp	Leu	Tyr	Leu	Pro	Glu	Glu	Val	Arg	Gly	Tyr	Ala
			115					120					125			
	Lys	Leu	Ala	Arg	Leu	Asp	Lys	Pro	Ile	Gly	Thr	Trp	Leu	Leu	Ala	Trp
		130					135					140				
	Pro	Cys	Met	Trp	Ser	Ile	Ala	Leu	Ala	Ala	Asp	Pro	Gly	Ser	Leu	Pro
20	145					150					155				160	
	Ser	Phe	Lys	Tyr	Met	Ala	Leu	Phe	Gly	Cys	Gly	Ala	Leu	Leu	Leu	Arg
				165					170						175	
	Gly	Ala	Gly	Cys	Thr	Ile	Asn	Asp	Leu	Leu	Asp	Gln	Asp	Ile	Asp	Thr
			180					185					190			
25	Lys	Val	Asp	Arg	Thr	Lys	Leu	Arg	Pro	Ile	Ala	Ser	Gly	Leu	Leu	Thr
			195					200					205			
	Pro	Phe	Gln	Gly	Ile	Gly	Phe	Leu	Gly	Leu	Gln	Leu	Leu	Leu	Gly	Leu
		210					215						220			
	Gly	Ile	Leu	Leu	Gln	Leu	Asn	Asn	Tyr	Ser	Arg	Val	Leu	Gly	Ala	Ser
30	225					230					235				240	
	Ser	Leu	Leu	Leu	Val	Phe	Ser	Tyr	Pro	Leu	Met	Lys	Arg	Phe	Thr	Phe
				245					250					255		
	Trp	Pro	Gln	Ala	Phe	Leu	Gly	Leu	Thr	Ile	Asn	Trp	Gly	Ala	Leu	Leu
			260					265					270			
35	Gly	Trp	Thr	Ala	Val	Lys	Gly	Ser	Ile	Ala	Pro	Ser	Ile	Val	Leu	Pro
			275					280					285			
	Leu	Tyr	Leu	Ser	Gly	Val	Cys	Trp	Thr	Leu	Val	Tyr	Asp	Thr	Ile	Tyr
		290					295					300				
	Ala	His	Gln	Asp	Lys	Glu	Asp	Asp	Val	Lys	Val	Gly	Val	Lys	Ser	Thr
40	305					310					315				320	
	Ala	Leu	Arg	Phe	Gly	Asp	Asn	Thr	Lys	Leu	Trp	Leu	Thr	Gly	Phe	Gly
				325					330					335		
	Thr	Ala	Ser	Ile	Gly	Phe	Leu	Ala	Leu	Ser	Gly	Phe	Ser	Ala	Asp	Leu
			340					345					350			
45	Gly	Trp	Gln	Tyr	Tyr	Ala	Ser	Leu	Ala	Ala	Ala	Ser	Gly	Gln	Leu	Gly
			355					360					365			

Trp Gln Ile Gly Thr Ala Asp Leu Ser Ser Gly Ala Asp Cys Ser Arg
 370 375 380
 Lys Phe Val Ser Asn Lys Trp Phe Gly Ala Ile Ile Phe Ser Gly Val
 385 390 395 400
 5 Val Leu Gly Arg Ser Phe Gln
 405

<210> 5
 <211> 1296
 10 <212> DNA
 <213> Arabidopsis sp

<400> 5
 atgtggcgaa gatctgttgt ttctcgttta tcttcaagaa tctctgtttc ttcttcgtta 60
 15 ccaaacccta gactgattcc ttggtcccg cgaattatgtg ccgttaatag cttctcccag 120
 cctccggtct cgacggaatc aactgctaag ttagggatca ctggtgtag atctgatgcc 180
 aatcgagttt ttgccactgc tactgcccgc gctacagcta cagctaccac cggtgagatt 240
 tcgtctagag ttgcggcttt ggctggatta gggcatcact acgctcgttg ttattgggag 300
 ctttctaaag ctaaaacttag tatgcttgtg gttgcaactt ctggaactgg gtatattctg 360
 20 ggtacgggaa atgctgcaat tagcttcccg gggctttgtt acacatgtgc aggaaccatg 420
 atgattgctg catctgctaa ttcttgaat cagatttttg agataagcaa tgattctaag 480
 atgaaaagaa cgatgctaag gccattgcct tcaggacgta ttagtgttcc acacgctgtt 540
 gcatgggcta ctattgctgg tgcttctggg gcttggttgt tggccagcaa gactaatatg 600
 ttggctgctg gacttgcata tgccaatctt gtactttatg cgtttgttta tactccgttg 660
 25 aagcaacttc accctatcaa tacatgggtt ggcgctgttg ttggtgctat cccacccttg 720
 cttgggtggg cggcagcgtc tggtcagatt tcatacaatt cgatgattct tccagctgct 780
 ctttactttt ggcagatacc tcattttatg gcccttgcaac atctctgccg caatgattat 840
 gcagctggag gttacaagat gttgtcactc tttgatccgt caggggaagag aatagcagca 900
 gtggctctaa ggaactgctt ttacatgata cctctcggtt tcatcgcccta tgactggggg 960
 30 ttaacctcaa gttggttttg cctcgaatca acacttctca cactagcaat cgctgcaaca 1020
 gcattttcat tctaccgaga cgggaccatg cataaagcaa ggaaaatgtt ccatgccagt 1080
 cttctcttcc ttctgtttt catgtctggg cttcttctac accgtgtctc taatgataat 1140
 cagcaacaac tcgtagaaga agccggatta acaaattctg tatctggtga agtcaaaact 1200
 cagagcgcaa agaaacgtgt ggctcaacct ccggtggctt atgcctctgc tgcaccgttt 1260
 35 cctttcctcc cagctccttc cttctactct ccatga 1296

<210> 6
 <211> 431
 <212> PRT
 40 <213> Arabidopsis sp

<400> 6
 Met Trp Arg Arg Ser Val Val Tyr Arg Phe Ser Ser Arg Ile Ser Val
 1 5 10 15
 45 Ser Ser Ser Leu Pro Asn Pro Arg Leu Ile Pro Trp Ser Arg Glu Leu
 20 25 30

Cys Ala Val Asn Ser Phe Ser Gln Pro Pro Val Ser Thr Glu Ser Thr
 35 40 45
 Ala Lys Leu Gly Ile Thr Gly Val Arg Ser Asp Ala Asn Arg Val Phe
 50 55 60
 5 Ala Thr Ala Thr Ala Ala Ala Thr Ala Thr Thr Thr Gly Glu Ile
 65 70 75 80
 Ser Ser Arg Val Ala Ala Leu Ala Gly Leu Gly His His Tyr Ala Arg
 85 90 95
 10 Cys Tyr Trp Glu Leu Ser Lys Ala Lys Leu Ser Met Leu Val Val Ala
 100 105 110
 Thr Ser Gly Thr Gly Tyr Ile Leu Gly Thr Gly Asn Ala Ala Ile Ser
 115 120 125
 Phe Pro Gly Leu Cys Tyr Thr Cys Ala Gly Thr Met Met Ile Ala Ala
 130 135 140
 15 Ser Ala Asn Ser Leu Asn Gln Ile Phe Glu Ile Ser Asn Asp Ser Lys
 145 150 155 160
 Met Lys Arg Thr Met Leu Arg Pro Leu Pro Ser Gly Arg Ile Ser Val
 165 170 175
 Pro His Ala Val Ala Trp Ala Thr Ile Ala Gly Ala Ser Gly Ala Cys
 180 185 190
 20 Leu Leu Ala Ser Lys Thr Asn Met Leu Ala Ala Gly Leu Ala Ser Ala
 195 200 205
 Asn Leu Val Leu Tyr Ala Phe Val Tyr Thr Pro Leu Lys Gln Leu His
 210 215 220
 25 Pro Ile Asn Thr Trp Val Gly Ala Val Val Gly Ala Ile Pro Pro Leu
 225 230 235 240
 Leu Gly Trp Ala Ala Ala Ser Gly Gln Ile Ser Tyr Asn Ser Met Ile
 245 250 255
 Leu Pro Ala Ala Leu Tyr Phe Trp Gln Ile Pro His Phe Met Ala Leu
 260 265 270
 30 Ala His Leu Cys Arg Asn Asp Tyr Ala Ala Gly Gly Tyr Lys Met Leu
 275 280 285
 Ser Leu Phe Asp Pro Ser Gly Lys Arg Ile Ala Ala Val Ala Leu Arg
 290 295 300
 35 Asn Cys Phe Tyr Met Ile Pro Leu Gly Phe Ile Ala Tyr Asp Trp Gly
 305 310 315 320
 Leu Thr Ser Ser Trp Phe Cys Leu Glu Ser Thr Leu Leu Thr Leu Ala
 325 330 335
 Ile Ala Ala Thr Ala Phe Ser Phe Tyr Arg Asp Arg Thr Met His Lys
 340 345 350
 40 Ala Arg Lys Met Phe His Ala Ser Leu Leu Phe Leu Pro Val Phe Met
 355 360 365
 Ser Gly Leu Leu Leu His Arg Val Ser Asn Asp Asn Gln Gln Gln Leu
 370 375 380
 45 Val Glu Glu Ala Gly Leu Thr Asn Ser Val Ser Gly Glu Val Lys Thr
 385 390 395 400

Gln Arg Arg Lys Lys Arg Val Ala Gln Pro Pro Val Ala Tyr Ala Ser
 405 410 415
 Ala Ala Pro Phe Pro Phe Leu Pro Ala Pro Ser Phe Tyr Ser Pro
 420 425 430

5
 <210> 7
 <211> 479
 <212> DNA
 <213> Arabidopsis sp

10
 <400> 7
 ggaaactccc ggagcacctg tttgcaggta ccgctaacct taatcgataa tttattttctc 60
 ttgtcaggaa ttatgtaagt ctgggtggaag gctcgcatac cattttttgca ttgccttttcg 120
 ctatgatcgg gtttactttg ggtgtgatga gaccaggcgt ggcttttatgg tatggcgaaa 180
 15 acccattttt atccaatgct gcattccctc ccgatgatgc gttctttcat tcctatacag 240
 gtatcatgct gataaaaactg ttactgggtac tgggtttgtat ggtatcagca agaagcgcgg 300
 cgatggcggt taaccgggtat ctgcacaggc attttgacgc gaagaaccgc cgtactgcca 360
 tccgtgaaat acctgcgggc gtcataatctg ccaacagtgc gctgggtgtt acgataggct 420
 gctgcgtggt attctgggtg gcctgttatt tcattaacac gatctgtttt tacctggcg 479

20
 <210> 8
 <211> 551
 <212> DNA
 <213> Arabidopsis sp

25
 <220>
 <221> misc_feature
 <222> (1)...(551)
 <223> n = A,T,C or G

30
 <400> 8
 ttgtggctta caccttaatg agcatacgcc agnccattac ggctcgtaa tcggcgccat 60
 ngccgngct gntgcaccgg tagtgggcta ctgcgccgtg accaatcagc ttgatctagc 120
 ggctcttatt ctgtttttta ttttactgtt ctggcaaatg ccgcattttt acgcgatttc 180
 35 cattttcagg ctaaaagact tttcagcggc ctgtattccg gtgctgcca tcattaaaga 240
 cctgcgctat accaaaatca gcatgctggt ttacgtgggc ttatttacac tggctgctat 300
 catgccgggc ctcttagggg atgccggtt gatttatggg atagcggcct taattttagg 360
 cttgtattgg ctttatattg ccatacaagg attcaagacc gccgatgatc aaaaatgggtc 420
 tcgtaagatg tttggatctt cgattttaat cattaccctc ttgtcggtta tgatgcttgt 480
 40 ttaaaacttac tgccctctga agtttatata tcgataattt cagcttaagg aggcttagtg 540
 gttaattcaa t 551

<210> 9
 <211> 297
 45 <212> PRT
 <213> Arabidopsis sp

<400> 9
 Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
 1 5 10 15
 5 Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
 20 25 30
 Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
 35 40 45
 Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
 10 50 55 60
 Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
 65 70 75 80
 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
 85 90 95
 15 Val Val Met Gly Asn Lys Val Val Ala Leu Leu Ala Thr Ala Val Glu
 100 105 110
 His Leu Val Thr Gly Glu Thr Met Glu Ile Thr Ser Ser Thr Glu Gln
 115 120 125
 Arg Tyr Ser Met Asp Tyr Tyr Met Gln Lys Thr Tyr Tyr Lys Thr Ala
 20 130 135 140
 Ser Leu Ile Ser Asn Ser Cys Lys Ala Val Ala Val Leu Thr Gly Gln
 145 150 155 160
 Thr Ala Glu Val Ala Val Leu Ala Phe Glu Tyr Gly Arg Asn Leu Gly
 165 170 175
 25 Leu Ala Phe Gln Leu Ile Asp Asp Ile Leu Asp Phe Thr Gly Thr Ser
 180 185 190
 Ala Ser Leu Gly Lys Gly Ser Leu Ser Asp Ile Arg His Gly Val Ile
 195 200 205
 Thr Ala Pro Ile Leu Phe Ala Met Glu Glu Phe Pro Gln Leu Arg Glu
 30 210 215 220
 Val Val Asp Gln Val Glu Lys Asp Pro Arg Asn Val Asp Ile Ala Leu
 225 230 235 240
 Glu Tyr Leu Gly Lys Ser Lys Gly Ile Gln Arg Ala Arg Glu Leu Ala
 245 250 255
 35 Met Glu His Ala Asn Leu Ala Ala Ala Ala Ile Gly Ser Leu Pro Glu
 260 265 270
 Thr Asp Asn Glu Asp Val Lys Arg Ser Arg Arg Ala Leu Ile Asp Leu
 275 280 285
 Thr His Arg Val Ile Thr Arg Asn Lys
 40 290 295

<210> 10

<211> 561

<212> DNA

45 <213> Arabidopsis sp

<400> 10
aagcgcatcc gtctcttctt acgattgccg ccagccgcat gtatggctgc ataaccgacc 60
gcccctatcc gctcgcggcc gcggtcgaat tcattcacac cgcgacgctg ctgcatgacg 120
acgtcgtcga tgaaagcgat ttgcgccgcg gccgcgaaag cgcgcataag gttttcggca 180
5 atcaggcgag cgtgctcgtc ggcgatttcc ttttctccc cgccttccag ctgatgggtg 240
aagacggctc gctcgcgcgc ctgcgcattc tctcggatgc ctccgccgtg atcgcgcagg 300
gcgaagtgat gcagctcggc accgcgcgca atcttgaaac caatatgagc cagtatctcg 360
atgtgatcag cgcgaagacc gccgcgctct ttgccgccgc ctgcgaaatc ggcccgggtga 420
tggcgaacgc gaaggcgga gatgctgccg cgatgtgcga atacggcatg aatctcggta 480
10 tgccttcca gatcatcgac gaccttctcg attacggcac cggcggccac gccgagcttg 540
gcaagaacac gggcgacgat t 561

<210> 11
<211> 966
15 <212> DNA
<213> Arabidopsis sp

<400> 11
atgggtacttg ccgaggttcc aaagcttgcc tctgctgctg agtacttctt caaaaggggt 60
20 gtgcaaggaa aacagtttct ttcaactatt ttgctgctga tggcgacagc tctgaatgta 120
cgcgttccag aagcattgat tggggaatca acagatatag tcacatcaga attacgcgta 180
aggcaacggg gtattgctga aatcactgaa atgatacacg tcgcaagtct actgcacgat 240
gatgtcttgg atgatgccga tacaaggcgt ggtgttggtt ccttaaagt tgtaatgggt 300
aacaagatgt cgggtattagc aggagacttc ttgctctccc gggcttggtg ggctctcgct 360
25 gctttaaaga acacagaggt tgtagcatta cttgcaactg ctgtagaaca tcttgttacc 420
ggtgaaacca tggaaataac tagttcaacc gagcagcgtt atagtatgga ctactacatg 480
cagaagacat attataagac agcatcgcta atctctaaca gctgcaaagc tgttgccgtt 540
ctcactggac aaacagcaga agttgccgtg ttagcttttg agtatgggag gaatctgggt 600
ttagcattcc aattaataga cgacattctt gatctcacgg gcacatctgc ctctctcgga 660
30 aagggatcgt tgtcagatat tcgccatgga gtcataacag cccaatcct ctttgccatg 720
gaagagtttc ctcaactacg cgaagttggt gatcaagttg aaaaagatcc taggaatggt 780
gacattgctt tagagtatct tgggaagagc aagggaatac agagggcaag agaattagcc 840
atggaacatg cgaatctagc agcagctgca atcgggtctc tacctgaaac agacaatgaa 900
gatgtcaaaa gatcgaggcg ggcacttatt gacttgaccc atagagtcac caccagaaac 960
35 aagtga 966

<210> 12
<211> 321
<212> PRT
40 <213> Arabidopsis sp

<400> 12
Met Val Leu Ala Glu Val Pro Lys Leu Ala Ser Ala Ala Glu Tyr Phe
1 5 10 15
45 Phe Lys Arg Gly Val Gln Gly Lys Gln Phe Arg Ser Thr Ile Leu Leu
20 25 30

Leu Met Ala Thr Ala Leu Asn Val Arg Val Pro Glu Ala Leu Ile Gly
 35 40 45
 Glu Ser Thr Asp Ile Val Thr Ser Glu Leu Arg Val Arg Gln Arg Gly
 50 55 60
 5 Ile Ala Glu Ile Thr Glu Met Ile His Val Ala Ser Leu Leu His Asp
 65 70 75 80
 Asp Val Leu Asp Asp Ala Asp Thr Arg Arg Gly Val Gly Ser Leu Asn
 85 90 95
 Val Val Met Gly Asn Lys Met Ser Val Leu Ala Gly Asp Phe Leu Leu
 10 100 105 110
 Ser Arg Ala Cys Gly Ala Leu Ala Ala Leu Lys Asn Thr Glu Val Val
 115 120 125
 Ala Leu Leu Ala Thr Ala Val Glu His Leu Val Thr Gly Glu Thr Met
 130 135 140
 15 Glu Ile Thr Ser Ser Thr Glu Gln Arg Tyr Ser Met Asp Tyr Tyr Met
 145 150 155 160
 Gln Lys Thr Tyr Tyr Lys Thr Ala Ser Leu Ile Ser Asn Ser Cys Lys
 165 170 175
 Ala Val Ala Val Leu Thr Gly Gln Thr Ala Glu Val Ala Val Leu Ala
 20 180 185 190
 Phe Glu Tyr Gly Arg Asn Leu Gly Leu Ala Phe Gln Leu Ile Asp Asp
 195 200 205
 Ile Leu Asp Phe Thr Gly Thr Ser Ala Ser Leu Gly Lys Gly Ser Leu
 210 215 220
 25 Ser Asp Ile Arg His Gly Val Ile Thr Ala Pro Ile Leu Phe Ala Met
 225 230 235 240
 Glu Glu Phe Pro Gln Leu Arg Glu Val Val Asp Gln Val Glu Lys Asp
 245 250 255
 Pro Arg Asn Val Asp Ile Ala Leu Glu Tyr Leu Gly Lys Ser Lys Gly
 30 260 265 270
 Ile Gln Arg Ala Arg Glu Leu Ala Met Glu His Ala Asn Leu Ala Ala
 275 280 285
 Ala Ala Ile Gly Ser Leu Pro Glu Thr Asp Asn Glu Asp Val Lys Arg
 290 295 300
 35 Ser Arg Arg Ala Leu Ile Asp Leu Thr His Arg Val Ile Thr Arg Asn
 305 310 315 320
 Lys

40 <210> 13

<211> 621

<212> DNA

<213> Arabidopsis sp

45 <400> 13

gctttctcct ttgctaattc ttgagctttc ttgatccac cgcgatttct aactatttca

60

atcgcttctt caagcgatcc aggctcacia aactcagact caatgatctc tcttagcctt 120
 ggctcattct ctagcgcgaa gatcactggc gccgttatgt tacctttggc taagtcatta 180
 gctgcaggct tacctaactg ctctgtggac tgagtgaagt ccagaatgac atcaactact 240
 tgaaaagata aaccgagatt cttcccgaac tgatacattt gctctgcgac cttgctttcg 300
 5 actttactga aaattgctgc tcctttgggtg cttgcagcta ctaatgaagc tgtctttag 360
 taactcttta gcatgtagtc atcaagcttg acatcacaat cgaataaact cgatgcttgc 420
 tttatctcac cgcttgcaaa atctttgatc acctgcaaaa agataaatca agattcgacac 480
 caaatgttct ttgtattgag tagcttcac taatctcaga aaggaatatt acctgactta 540
 tgagcttaat gacttcaagg ttttcgagat ttgtaagtac catgatgctt gagcaacatg 600
 10 aaatccccag ctaatacagc t 621

<210> 14

<211> 741

<212> DNA

15 <213> Arabidopsis sp

<400> 14
 ggtgagtttt gttaatatgt atgagattca tctatttttg tcataaaatt gtttggtttg 60
 gtttaaaact tgtgtataat tgcaggaaag gaaacagttc atgagctttt cggcacaaga 120
 20 gtagcgggtg tagctggaga tttcatgttt gctcaagcgt catggtactt agcaaatctc 180
 gagaatcttg aagttattaa gctcatcagt cagggtactta gttactctta cattgttttt 240
 ctatgagggt gagctatgaa tctcatttcg ttgaataatg ctgtgcctca aacttttttt 300
 catgttttca ggtgatcaaa gactttgcaa gcggagagat aaagcaggcg tccagcttat 360
 ttgactgcga caccaagctc gacgagtact tactcaaaag tttctacaag acagcctctt 420
 25 tagtggtctg gagcaccaaa ggagctgcca ttttcagcag agttgagcct gatgtgacag 480
 aacaaatgta cgagtttggg aagaatctcg gtctctcttt ccagatagtt gatgatattt 540
 tggatttcac tcagtgcaca gagcagctcg ggaagccagc agggagtgat ttggctaaag 600
 gtaacttaac agcacctgtg attttcgctc tggagagggg gccaaaggcta agagagatca 660
 ttgagtcaaa gttctgtgag gcgggttctc tggaagaagc gattgaagcg gtgacaaaag 720
 30 gtgggggggat taagagagca c 741

<210> 15

<211> 1087

<212> DNA

35 <213> Arabidopsis sp

<400> 15
 cctcttcagc caatccagag gaagaagaga caacttttta tctttcgtca agagtctcgg 60
 aaaacgcacg gttttatgct ctctcttctg ccctcacctc acaagacgca gggcacatga 120
 40 ttcaaccaga gggaaaaagc aacgataaca actctgcttt tgatttcaag ctgtatatga 180
 tccgcaaagc cgagtctgta aatgcggctc tcgacgtttc cgtaccgctt ctgaaacccc 240
 ttacgatcca agaagcggtc aggtactctt tgctagccgg cggaacacgt gtgaggcctc 300
 tgctctgcat tgccgcttgt gagcttgttg gggcgacga ggctactgcc atgtcagccg 360
 cttgcgcggg cgagatgatc cacacaagct ctctcattca tgacgatctt ccgtgcatgg 420
 45 acaatgccga cctccgtaga ggcaagccca ccaatcacia ggtatgttgt ttaattatat 480
 gaaggctcag agataatgct gaactagtgt tgaaccaatt tttgctcaa caaggtatat 540

ggagaagaca tggcgggtttt ggcaggtgat gcactccttg cattggcggtt tgagcacatg 600
 acggttgtgt cgagtgggtt ggtcgctccc gagaagatga ttcgcgccgt ggttgagctg 660
 gccagggcca tagggactac agggctagtt gctggacaaa tgatagacct agccagcgaa 720
 agactgaatc cagacaaggt tggattggag catctagagt tcatccatct ccacaaaacg 780
 5 gcggcattgt tggaggcagc ggcagtttta ggggttataa tgggaggtgg aacagaggaa 840
 gaaatcgaaa agcttagaaa gtatgctagg tgtattggac tactgtttca ggttgttgat 900
 gacattctcg acgtaacaaa atctactgag gaattgggta agacagccgg aaaagacgta 960
 atggccggaa agctgacgta tccaaggctg atagggttgg agggatccag ggaagttgca 1020
 gagcacctga ggagagaagc agaggaaaag cttaaagggt ttgatccaag tcaggcggcg 1080
 10 cctctgg 1087

<210> 16

<211> 1164

<212> DNA

15 <213> Arabidopsis sp

<400> 16

atgacttcga ttctcaacac tgtctccacc atccactctt ccagagttac ctccgtcgat 60
 cgagtcggag tcctctctct tcggaattcg gattccgttg agttcactcg ccggcggttct 120
 20 ggtttctcga cgttgatcta cgaatcacc gggcgagat ttgttggtgc tgccggcggag 180
 actgatactg ataaagttaa atctcagaca cctgacaagg caccagccgg tggttcaagc 240
 attaaccagc ttctcggtat caaaggagca tctcaagaaa ctaataaatg gaagattcgt 300
 cttcagctta caaaaccagt cacttggcct ccactgggtt ggggagtcgt ctgtggtgct 360
 gctgcttcag ggaactttca ttggaccca gaggatgttg ctaagtogat tctttgcatg 420
 25 atgatgtctg gtccttgtct tactggctat acacagacaa tcaacgactg gtatgataga 480
 gatatcgacg caattaatga gccatctcgt ccaattccat ctggagcaat atcagagcca 540
 gaggttatta cacaagtctg ggtgctatta ttgggaggtc ttggtattgc tggaatatta 600
 gatgtgtggg cagggcatac cactcccact gtcttctatc ttgctttggg aggatcattg 660
 ctatcttata tatactctgc tccactctt aagctaaaac aaaatggatg ggttggaat 720
 30 tttgacttg gagcaagcta tattagtttg ccattggtgg ctggccaagc attgtttggc 780
 actcttacgc cagatgttgt gtcttaaca ctctgtaca gcatagctgg gtttaggaata 840
 gccattgtta acgacttcaa aagtgttgaa ggagatagag cattaggact tcagtctctc 900
 ccagtagctt ttggcaccga aactgcaaaa tggatatgct ttggtgctat agacattact 960
 cagctttctg ttgccggata tctattagca tctgggaaac cttattatgc gttggcggtg 1020
 35 gttgctttga tcattcctca gattgtgttc cagtttaaat actttctcaa ggaccctgtc 1080
 aaatacgacg tcaagtacca ggcaagcgcg cagccattct tgggtgctcg aatatttgta 1140
 acggcattag catcgcaaca ctga 1164

<210> 17

40 <211> 387

<212> PRT

<213> Arabidopsis sp

<400> 17

45 Met Thr Ser Ile Leu Asn Thr Val Ser Thr Ile His Ser Ser Arg Val

1

5

10

15

	Thr	Ser	Val	Asp	Arg	Val	Gly	Val	Leu	Ser	Leu	Arg	Asn	Ser	Asp	Ser	
				20					25					30			
	Val	Glu	Phe	Thr	Arg	Arg	Arg	Ser	Gly	Phe	Ser	Thr	Leu	Ile	Tyr	Glu	
			35					40					45				
5	Ser	Pro	Gly	Arg	Arg	Phe	Val	Val	Arg	Ala	Ala	Glu	Thr	Asp	Thr	Asp	
		50					55					60					
	Lys	Val	Lys	Ser	Gln	Thr	Pro	Asp	Lys	Ala	Pro	Ala	Gly	Gly	Ser	Ser	
	65					70				75					80		
	Ile	Asn	Gln	Leu	Leu	Gly	Ile	Lys	Gly	Ala	Ser	Gln	Glu	Thr	Asn	Lys	
10				85						90					95		
	Trp	Lys	Ile	Arg	Leu	Gln	Leu	Thr	Lys	Pro	Val	Thr	Trp	Pro	Pro	Leu	
				100					105					110			
	Val	Trp	Gly	Val	Val	Cys	Gly	Ala	Ala	Ala	Ser	Gly	Asn	Phe	His	Trp	
			115					120					125				
15	Thr	Pro	Glu	Asp	Val	Ala	Lys	Ser	Ile	Leu	Cys	Met	Met	Met	Ser	Gly	
		130					135					140					
	Pro	Cys	Leu	Thr	Gly	Tyr	Thr	Gln	Thr	Ile	Asn	Asp	Trp	Tyr	Asp	Arg	
	145					150					155					160	
	Asp	Ile	Asp	Ala	Ile	Asn	Glu	Pro	Tyr	Arg	Pro	Ile	Pro	Ser	Gly	Ala	
20					165					170					175		
	Ile	Ser	Glu	Pro	Glu	Val	Ile	Thr	Gln	Val	Trp	Val	Leu	Leu	Leu	Gly	
				180					185					190			
	Gly	Leu	Gly	Ile	Ala	Gly	Ile	Leu	Asp	Val	Trp	Ala	Gly	His	Thr	Thr	
			195					200					205				
25	Pro	Thr	Val	Phe	Tyr	Leu	Ala	Leu	Gly	Gly	Ser	Leu	Leu	Ser	Tyr	Ile	
		210					215					220					
	Tyr	Ser	Ala	Pro	Pro	Leu	Lys	Leu	Lys	Gln	Asn	Gly	Trp	Val	Gly	Asn	
	225					230					235					240	
	Phe	Ala	Leu	Gly	Ala	Ser	Tyr	Ile	Ser	Leu	Pro	Trp	Trp	Ala	Gly	Gln	
30					245					250					255		
	Ala	Leu	Phe	Gly	Thr	Leu	Thr	Pro	Asp	Val	Val	Val	Leu	Thr	Leu	Leu	
			260					265					270				
	Tyr	Ser	Ile	Ala	Gly	Leu	Gly	Ile	Ala	Ile	Val	Asn	Asp	Phe	Lys	Ser	
		275						280					285				
35	Val	Glu	Gly	Asp	Arg	Ala	Leu	Gly	Leu	Gln	Ser	Leu	Pro	Val	Ala	Phe	
		290					295					300					
	Gly	Thr	Glu	Thr	Ala	Lys	Trp	Ile	Cys	Val	Gly	Ala	Ile	Asp	Ile	Thr	
	305					310					315				320		
	Gln	Leu	Ser	Val	Ala	Gly	Tyr	Leu	Leu	Ala	Ser	Gly	Lys	Pro	Tyr	Tyr	
40					325					330					335		
	Ala	Leu	Ala	Leu	Val	Ala	Leu	Ile	Ile	Pro	Gln	Ile	Val	Phe	Gln	Phe	
				340					345					350			
	Lys	Tyr	Phe	Leu	Lys	Asp	Pro	Val	Lys	Tyr	Asp	Val	Lys	Tyr	Gln	Ala	
		355					360					365					
45	Ser	Ala	Gln	Pro	Phe	Leu	Val	Leu	Gly	Ile	Phe	Val	Thr	Ala	Leu	Ala	
		370					375					380					

Ser Gln His

385

<210> 18

5 <211> 981

<212> DNA

<213> Arabidopsis sp

<400> 18

10	atgtttgttta gtggttcagc gatcccattha agcagcttct gctctcttcc ggagaaaccc	60
	cacactcttc ctatgaaact ctctcccgct gcaatccgat ctctatctc atctgccccg	120
	gggtcggtga acttcgatct gaggaagtat tggacgactc tgatcaccga gatcaaccag	180
	aagctggatg aggccatacc ggtcaagcac cctgcgggga tctacgagge tatgagatac	240
	tctgtactcg cacaaggcgc caagcgtgcc cctcctgtga tgtgtgtggc ggcctgcgag	300
15	ctcttcgggtg gcgatcgccct cgccgctttc cccaccgcct gtgccctaga aatggtgcac	360
	gcggcttctgt tgatacacga cgacctcccc tgtatggacg acgatcctgt gcgcagagga	420
	aagccatcta accacactgt ctacggctct ggcatggcca ttctcgccgg tgacgccctc	480
	ttcccactcg ccttccagca cattgtctcc cacacgcctc ctgaccttgt tccccgagcc	540
	accatcctca gactcatcac tgagattgcc cgcactgtcg gctccactgg tatggctgca	600
20	ggccagtagc tgcacctga aggaggtccc ttctctcttt cctttgttca ggagaagaaa	660
	ttcggagcca tgggtgaatg ctctgccgtg tgcggtggcc tattgggagg tgccactgag	720
	gatgagctcc agagtctccg aaggtagggg agagccgtcg ggatgctgta tcagggtggtc	780
	gatgacatca ccgaggacaa gaagaagagc tatgatggtg gagcagagaa gggaatgatg	840
	gaaatggcgg aagagctcaa ggagaaggcg aagaaggagc ttcaagtgtt tgacaacaag	900
25	tatggaggag gagacacact tgttcctctc tacaccttcg ttgactacgc tgctcatcga	960
	cattttcttc tccccctctg a	981

<210> 19

<211> 245

30 <212> DNA

<213> Glycine sp

<400> 19

35	gcaacatctg ggactgggtt tgtcttgggg agtggttagtg ctgttgatct ttccggcactt	60
	tcttgcaactt gcttgggtac catgatggtt gctgcatctg ctaactcttt gaatcaggtg	120
	tttgagatca ataagatgac taaaatgaag agaacaagtc gcaggccact accctcagga	180
	cgcatacaaa tacctcatgc agttggctgg gcatcctctg ttggattagc tggtagcggt	240
	ctact	245

40 <210> 20

<211> 253

<212> DNA

<213> Glycine sp

45 <400> 20

attggctttc caagatcatt ggggttttctt gttgcattca tgaccttcta ctccttgggt	60
--	----

ttggcattgt ccaaggatat acctgacgtt gaaggagata aagagcacgg cattgattct 120
 tttgcagtac gtctaggtca gaaacgggca ttttggattt gcgtttcctt ttttgaaatg 180
 gctttcggag ttggtatcct ggccggagca tcatgctcac acttttggac taaaattttc 240
 acgggtatgg gaa 253

5

<210> 21
 <211> 275
 <212> DNA
 <213> Glycine sp

10

<400> 21
 tgatcttcta ctctctgggt atggcattgt ccaaggatat atctgacgtt aaaggagata 60
 aagcatacgg catcgatact ttagcgatac gtttgggtca aaaatgggta ttttggattt 120
 gcattatcct ttttgaaatg gcttttggag ttgccctcctt ggcaggagca acatcttctt 180
 15 acctttggat taaaattgtc acgggtcttg gacatgctat tcttgcttca attctcttct 240
 accaagccaa atctatatac ttgagcaaca aagtt 275

<210> 22
 <211> 299
 20 <212> DNA
 <213> Glycine sp

<220>
 <221> misc_feature
 25 <222> (1)...(299)
 <223> n = A,T,C or G

<400> 22
 ccanaatang tncatcttng aaagacaatt ggcctcttca acacacaagt ctgcatgtga 60
 30 agaagaggcc aattgtcttt ccaagatcac ttatngtggc tattgtaatc atgaacttct 120
 tctttgtggg tatggcattg gcaaaggata tacctanctg ttgaaggaga taaaatatac 180
 ggcattgata cttttgcaat acgtataggt caaaaacaag tattttggat ttgtattttc 240
 ctttttgaaa ggctttcggg gtttccctag tggcaggagc aacatcttct agccttggg 299

35

<210> 23
 <211> 767
 <212> DNA
 <213> Glycine sp

40

<400> 23
 gtggaggctg tggttgctgc cctgtttatg aatatttata ttgttgggtt gaatcaattg 60
 tctgatgttg aaatagacaa gataaacaag ccgtatcttc cattagcatc tggggaatat 120
 tcctttgaaa ctggtgtcac tattgttgca tctttttcaa ttctgagttt ttggcttggc 180
 tgggtttag gttcatggcc attatatttg gccctttttg taagctttgt gctaggaact 240
 45 gcttattcaa tcaatgtgcc tctgttgaga tggaagaggt ttgcagtgc tgcagcgatg 300
 tgcattctag ctgttcgggc agtaatagtt caacttgcac ttttcttca catgcagact 360

catgtgtaca agaggccacc tgtcttttca agaccattga tttttgctac tgcattcatg 420
 agcttcttct ctgtagttat agcactgttt aaggatatac ctgacattga aggagataaa 480
 gtatttgga tccaatcttt ttcagtgtgt ttaggtcaga agccgggtgt ctggacttgt 540
 gttacccttc ttgaaatagc ttatggagtc gccctcctgg tgggagctgc atctccttgt 600
 5 ctttggagca aaattttcac gggctctggga cacgctgtgc tggcttcaat tctctggttt 660
 catgccaaat ctgtagattt gaaaagcaaa gcttcgataa catccttcta tatgtttatt 720
 tggaagctat tttatgcaga atacttactc attccttttg ttagatg 767

<210> 24
 10 <211> 255
 <212> PRT
 <213> Glycine sp

<400> 24
 15 Val Glu Ala Val Val Ala Ala Leu Phe Met Asn Ile Tyr Ile Val Gly
 1 5 10 15
 Leu Asn Gln Leu Ser Asp Val Glu Ile Asp Lys Ile Asn Lys Pro Tyr
 20 25 30
 Leu Pro Leu Ala Ser Gly Glu Tyr Ser Phe Glu Thr Gly Val Thr Ile
 20 35 40 45
 Val Ala Ser Phe Ser Ile Leu Ser Phe Trp Leu Gly Trp Val Val Gly
 50 55 60
 Ser Trp Pro Leu Phe Trp Ala Leu Phe Val Ser Phe Val Leu Gly Thr
 65 70 75 80
 25 Ala Tyr Ser Ile Asn Val Pro Leu Leu Arg Trp Lys Arg Phe Ala Val
 85 90 95
 Leu Ala Ala Met Cys Ile Leu Ala Val Arg Ala Val Ile Val Gln Leu
 100 105 110
 Ala Phe Phe Leu His Met Gln Thr His Val Tyr Lys Arg Pro Pro Val
 30 115 120 125
 Phe Ser Arg Pro Leu Ile Phe Ala Thr Ala Phe Met Ser Phe Phe Ser
 130 135 140
 Val Val Ile Ala Leu Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Lys
 145 150 155 160
 35 Val Phe Gly Ile Gln Ser Phe Ser Val Cys Leu Gly Gln Lys Pro Val
 165 170 175
 Phe Trp Thr Cys Val Thr Leu Leu Glu Ile Ala Tyr Gly Val Ala Leu
 180 185 190
 Leu Val Gly Ala Ala Ser Pro Cys Leu Trp Ser Lys Ile Phe Thr Gly
 40 195 200 205
 Leu Gly His Ala Val Leu Ala Ser Ile Leu Trp Phe His Ala Lys Ser
 210 215 220
 Val Asp Leu Lys Ser Lys Ala Ser Ile Thr Ser Phe Tyr Met Phe Ile
 225 230 235 240
 45 Trp Lys Leu Phe Tyr Ala Glu Tyr Leu Leu Ile Pro Phe Val Arg
 245 250 255

<210> 25
 <211> 360
 <212> DNA
 5 <213> Zea sp

<220>
 <221> misc_feature
 <222> (1)...(360)
 10 <223> n = A,T,C or G

<400> 25
 ggcgtcttca cttgttcttg tcttctcgta tcccctgatg aagaggttca ctttttgcc 60
 tcaggcttat cttggcctga cattcaactg gggagcttta ctaggggtgg ctgctattaa 120
 15 ggaaagcata gaccctgcaa atcatccttc cattgtatac agctgggtatt tggtggacgc 180
 tgggtgatga tactatataat gcgcacagc tggttcgcta tccctacttt catattaatc 240
 cttgatgaag tggccatttc atgttgctgc ggtgggtctta tacttgcata tctccatgca 300
 tctcaggaca aagangatga cctgaaagta ggagtccaag tccacagctt aagatttggg 360

20 <210> 26
 <211> 299
 <212> DNA
 <213> Zea sp

25 <220>
 <221> misc_feature
 <222> (1)...(299)
 <223> n = A,T,C or G

30 <400> 26
 gatggttgca gcatctgcaa ataccctcaa ccagggtgtt gngataaaaa atgatgctaa 60
 aatgaaaagg acaatgcgtg cccctgcca tctggctgca ttagtcctgc acatgctgcg 120
 atgtgggcta caagtgttg agttgcagga acagctttgt tggcctggaa ggctaattggc 180
 ttggcagctg ggcttgagc ttctaattctt gttctgtatg catttgtgta tacgccgttg 240
 35 aagcaaatac accctgttaa tacatgggtt ggggcagtcg ttggtgccat cccaccact 299

<210> 27
 <211> 255
 <212> DNA
 40 <213> Zea sp

<220>
 <221> misc_feature
 <222> (1)...(255)
 45 <223> n = A,T,C or G

<400> 27
 anacttgc atctccatgc ntctcaggac aaagangatg acctgaaagt aggtgtcaag 60
 tccacagcat taagatttgg agatttgacc nnatactgna tcagtggctt tggcgcgga 120
 tgcttcggca gcttagcact cagtggttac aatgctgacc ttggttgggtg tttagtgtga 180
 5 tgcttgagcg aagaatggta tngtttttac ttgatattga ctccagacct gaaatcatgt 240
 tggacagggg ggccc 255

<210> 28
 <211> 257
 10 <212> DNA
 <213> Zea sp

<400> 28
 attgaagggg ataggactct ggggcttcag tcacttcctg ttgcttttgg gatggaaact 60
 15 gcaaaatgga tttgtgttgg agcaattgat atcactcaat tatctgttgc aggttaccta 120
 ttgagcaccg gtaagctgta ttatgccctg gtgttgcttg ggctaacaat tcctcagggtg 180
 ttctttcagt tccagtactt cctgaaggac cctgtgaagt atgatgtcaa ataccaggca 240
 agcgcacaa cattctt 257

20 <210> 29
 <211> 368
 <212> DNA
 <213> Zea sp

25 <400> 29
 atccagttgc aaataataat ggcgttcttc tctgttgtaa tagcactatt caaggatata 60
 cctgacatcg aaggggaccg catattcggg atccgatcct tcagcgctcg gttagggcaa 120
 aagaaggtct tttggatctg cgttggtctg cttgagatgg cctacagcgt tgcgatactg 180
 atgggagcta cctcttcctg tttgtggagc aaaacagcaa ccatcgctgg ccattccata 240
 30 cttgccgcga tcctatggag ctgcgcgcga tcggtggact tgacgagcaa agccgcaata 300
 acgtccttct acatgttcat ctggaagctg ttctacgcgg agtacctgct catccctctg 360
 gtgcggtg 368

<210> 30
 35 <211> 122
 <212> PRT
 <213> Zea sp

<400> 30
 40 Ile Gln Leu Gln Ile Ile Met Ala Phe Phe Ser Val Val Ile Ala Leu
 1 5 10 15
 Phe Lys Asp Ile Pro Asp Ile Glu Gly Asp Arg Ile Phe Gly Ile Arg
 20 25 30
 Ser Phe Ser Val Arg Leu Gly Gln Lys Lys Val Phe Trp Ile Cys Val
 45 35 40 45
 Gly Leu Leu Glu Met Ala Tyr Ser Val Ala Ile Leu Met Gly Ala Thr

50 55 60
 Ser Ser Cys Leu Trp Ser Lys Thr Ala Thr Ile Ala Gly His Ser Ile
 65 70 75 80
 Leu Ala Ala Ile Leu Trp Ser Cys Ala Arg Ser Val Asp Leu Thr Ser
 5 85 90 95
 Lys Ala Ala Ile Thr Ser Phe Tyr Met Phe Ile Trp Lys Leu Phe Tyr
 100 105 110
 Ala Glu Tyr Leu Leu Ile Pro Leu Val Arg
 115 120
 10
 <210> 31
 <211> 278
 <212> DNA
 <213> Zea sp
 15
 <400> 31
 tattcagcac cacctctcaa gctcaagcag aatggatgga ttgggaactt cgctctgggt 60
 gcgagttaca tcagcttgcc ctgggtgggct ggccaggcgt tatttggaac tcttacacca 120
 gatattcattg tcttgactac tttgtacagc atagctgggc tagggattgc tattgtaaat 180
 20 gatttcaaga gtattgaagg ggataggact ctggggccttc agtcacttcc tggtgctttt 240
 gggatggaaa ctgcaaaatg gatttgtgtt ggagcaat 278

 <210> 32
 <211> 292
 25 <212> PRT
 <213> Synechocystis sp

 <400> 32
 Met Val Ala Gln Thr Pro Ser Ser Pro Pro Leu Trp Leu Thr Ile Ile
 30 1 5 10 15
 Tyr Leu Leu Arg Trp His Lys Pro Ala Gly Arg Leu Ile Leu Met Ile
 20 25 30
 Pro Ala Leu Trp Ala Val Cys Leu Ala Ala Gln Gly Leu Pro Pro Leu
 35 35 40 45
 35 Pro Leu Leu Gly Thr Ile Ala Leu Gly Thr Leu Ala Thr Ser Gly Leu
 50 55 60
 Gly Cys Val Val Asn Asp Leu Trp Asp Arg Asp Ile Asp Pro Gln Val
 65 70 75 80
 Glu Arg Thr Lys Gln Arg Pro Leu Ala Ala Arg Ala Leu Ser Val Gln
 40 85 90 95
 Val Gly Ile Gly Val Ala Leu Val Ala Leu Leu Cys Ala Ala Gly Leu
 100 105 110
 Ala Phe Tyr Leu Thr Pro Leu Ser Phe Trp Leu Cys Val Ala Ala Val
 115 120 125
 45 Pro Val Ile Val Ala Tyr Pro Gly Ala Lys Arg Val Phe Pro Val Pro
 130 135 140

Gln Leu Val Leu Ser Ile Ala Trp Gly Phe Ala Val Leu Ile Ser Trp
 145 150 155 160
 Ser Ala Val Thr Gly Asp Leu Thr Asp Ala Thr Trp Val Leu Trp Gly
 165 170 175
 5 Ala Thr Val Phe Trp Thr Leu Gly Phe Asp Thr Val Tyr Ala Met Ala
 180 185 190
 Asp Arg Glu Asp Asp Arg Arg Ile Gly Val Asn Ser Ser Ala Leu Phe
 195 200 205
 Phe Gly Gln Tyr Val Gly Glu Ala Val Gly Ile Phe Phe Ala Leu Thr
 10 210 215 220
 Ile Gly Cys Leu Phe Tyr Leu Gly Met Ile Leu Met Leu Asn Pro Leu
 225 230 235 240
 Tyr Trp Leu Ser Leu Ala Ile Ala Ile Val Gly Trp Val Ile Gln Tyr
 245 250 255
 15 Ile Gln Leu Ser Ala Pro Thr Pro Glu Pro Lys Leu Tyr Gly Gln Ile
 260 265 270
 Phe Gly Gln Asn Val Ile Ile Gly Phe Val Leu Leu Ala Gly Met Leu
 275 280 285
 Leu Gly Trp Leu
 20 290

 <210> 33
 <211> 316
 <212> PRT
 25 <213> Synechocystis sp

 <400> 33
 Met Val Thr Ser Thr Lys Ile His Arg Gln His Asp Ser Met Gly Ala
 1 5 10 15
 30 Val Cys Lys Ser Tyr Tyr Gln Leu Thr Lys Pro Arg Ile Ile Pro Leu
 20 25 30
 Leu Leu Ile Thr Thr Ala Ala Ser Met Trp Ile Ala Ser Glu Gly Arg
 35 40 45
 Val Asp Leu Pro Lys Leu Leu Ile Thr Leu Leu Gly Gly Thr Leu Ala
 35 50 55 60
 Ala Ala Ser Ala Gln Thr Leu Asn Cys Ile Tyr Asp Gln Asp Ile Asp
 65 70 75 80
 Tyr Glu Met Leu Arg Thr Arg Ala Arg Pro Ile Pro Ala Gly Lys Val
 85 90 95
 40 Gln Pro Arg His Ala Leu Ile Phe Ala Leu Ala Leu Gly Val Leu Ser
 100 105 110
 Phe Ala Leu Leu Ala Thr Phe Val Asn Val Leu Ser Gly Cys Leu Ala
 115 120 125
 Leu Ser Gly Ile Val Phe Tyr Met Leu Val Tyr Thr His Trp Leu Lys
 45 130 135 140
 Arg His Thr Ala Gln Asn Ile Val Ile Gly Gly Ala Ala Gly Ser Ile

145 150 155 160
 Pro Pro Leu Val Gly Trp Ala Ala Val Thr Gly Asp Leu Ser Trp Thr
 165 170 175
 Pro Trp Val Leu Phe Ala Leu Ile Phe Leu Trp Thr Pro Pro His Phe
 5 180 185 190
 Trp Ala Leu Ala Leu Met Ile Lys Asp Asp Tyr Ala Gln Val Asn Val
 195 200 205
 Pro Met Leu Pro Val Ile Ala Gly Glu Glu Lys Thr Val Ser Gln Ile
 210 215 220
 10 Trp Tyr Tyr Ser Leu Leu Val Val Pro Phe Ser Leu Leu Leu Val Tyr
 225 230 235 240
 Pro Leu His Gln Leu Gly Ile Leu Tyr Leu Ala Ile Ala Ile Ile Leu
 245 250 255
 Gly Gly Gln Phe Leu Val Lys Ala Trp Gln Leu Lys Gln Ala Pro Gly
 15 260 265 270
 Asp Arg Asp Leu Ala Arg Gly Leu Phe Lys Phe Ser Ile Phe Tyr Leu
 275 280 285
 Met Leu Leu Cys Leu Ala Met Val Ile Asp Ser Leu Pro Val Thr His
 290 295 300
 20 Gln Leu Val Ala Gln Met Gly Thr Leu Leu Leu Gly
 305 310 315

 <210> 34
 <211> 324
 25 <212> PRT
 <213> Synechocystis sp

 <400> 34
 Met Ser Asp Thr Gln Asn Thr Gly Gln Asn Gln Ala Lys Ala Arg Gln
 30 1 5 10 15
 Leu Leu Gly Met Lys Gly Ala Ala Pro Gly Glu Ser Ser Ile Trp Lys
 20 25 30
 Ile Arg Leu Gln Leu Met Lys Pro Ile Thr Trp Ile Pro Leu Ile Trp
 35 40 45
 35 Gly Val Val Cys Gly Ala Ala Ser Ser Gly Gly Tyr Ile Trp Ser Val
 50 55 60
 Glu Asp Phe Leu Lys Ala Leu Thr Cys Met Leu Leu Ser Gly Pro Leu
 65 70 75 80
 Met Thr Gly Tyr Thr Gln Thr Leu Asn Asp Phe Tyr Asp Arg Asp Ile
 40 85 90 95
 Asp Ala Ile Asn Glu Pro Tyr Arg Pro Ile Pro Ser Gly Ala Ile Ser
 100 105 110
 Val Pro Gln Val Val Thr Gln Ile Leu Ile Leu Leu Val Ala Gly Ile
 115 120 125
 45 Gly Val Ala Tyr Gly Leu Asp Val Trp Ala Gln His Asp Phe Pro Ile
 130 135 140

Met Met Val Leu Thr Leu Gly Gly Ala Phe Val Ala Tyr Ile Tyr Ser
 145 150 155 160
 Ala Pro Pro Leu Lys Leu Lys Gln Asn Gly Trp Leu Gly Asn Tyr Ala
 165 170 175
 5 Leu Gly Ala Ser Tyr Ile Ala Leu Pro Trp Trp Ala Gly His Ala Leu
 180 185 190
 Phe Gly Thr Leu Asn Pro Thr Ile Met Val Leu Thr Leu Ile Tyr Ser
 195 200 205
 Leu Ala Gly Leu Gly Ile Ala Val Val Asn Asp Phe Lys Ser Val Glu
 10 210 215 220
 Gly Asp Arg Gln Leu Gly Leu Lys Ser Leu Pro Val Met Phe Gly Ile
 225 230 235 240
 Gly Thr Ala Ala Trp Ile Cys Val Ile Met Ile Asp Val Phe Gln Ala
 245 250 255
 15 Gly Ile Ala Gly Tyr Leu Ile Tyr Val His Gln Gln Leu Tyr Ala Thr
 260 265 270
 Ile Val Leu Leu Leu Leu Ile Pro Gln Ile Thr Phe Gln Asp Met Tyr
 275 280 285
 Phe Leu Arg Asn Pro Leu Glu Asn Asp Val Lys Tyr Gln Ala Ser Ala
 20 290 295 300
 Gln Pro Phe Leu Val Phe Gly Met Leu Ala Thr Gly Leu Ala Leu Gly
 305 310 315 320
 His Ala Gly Ile

 25
 <210> 35
 <211> 307
 <212> PRT
 <213> Synechocystis sp

 30
 <400> 35
 Met Thr Glu Ser Ser Pro Leu Ala Pro Ser Thr Ala Pro Ala Thr Arg
 1 5 10 15
 Lys Leu Trp Leu Ala Ala Ile Lys Pro Pro Met Tyr Thr Val Ala Val
 35 20 25 30
 Val Pro Ile Thr Val Gly Ser Ala Val Ala Tyr Gly Leu Thr Gly Gln
 35 40 45
 Trp His Gly Asp Val Phe Thr Ile Phe Leu Leu Ser Ala Ile Ala Ile
 50 55 60
 40 Ile Ala Trp Ile Asn Leu Ser Asn Asp Val Phe Asp Ser Asp Thr Gly
 65 70 75 80
 Ile Asp Val Arg Lys Ala His Ser Val Val Asn Leu Thr Gly Asn Arg
 85 90 95
 Asn Leu Val Phe Leu Ile Ser Asn Phe Phe Leu Leu Ala Gly Val Leu
 45 100 105 110
 Gly Leu Met Ser Met Ser Trp Arg Ala Gln Asp Trp Thr Val Leu Glu

115 120 125
 Leu Ile Gly Val Ala Ile Phe Leu Gly Tyr Thr Tyr Gln Gly Pro Pro
 130 135 140
 Phe Arg Leu Gly Tyr Leu Gly Leu Gly Glu Leu Ile Cys Leu Ile Thr
 5 145 150 155 160
 Phe Gly Pro Leu Ala Ile Ala Ala Ala Tyr Tyr Ser Gln Ser Gln Ser
 165 170 175
 Phe Ser Trp Asn Leu Leu Thr Pro Ser Val Phe Val Gly Ile Ser Thr
 180 185 190
 10 Ala Ile Ile Leu Phe Cys Ser His Phe His Gln Val Glu Asp Asp Leu
 195 200 205
 Ala Ala Gly Lys Lys Ser Pro Ile Val Arg Leu Gly Thr Lys Leu Gly
 210 215 220
 Ser Gln Val Leu Thr Leu Ser Val Val Ser Leu Tyr Leu Ile Thr Ala
 15 225 230 235 240
 Ile Gly Val Leu Cys His Gln Ala Pro Trp Gln Thr Leu Leu Ile Ile
 245 250 255
 Ala Ser Leu Pro Trp Ala Val Gln Leu Ile Arg His Val Gly Gln Tyr
 260 265 270
 20 His Asp Gln Pro Glu Gln Val Ser Asn Cys Lys Phe Ile Ala Val Asn
 275 280 285
 Leu His Phe Phe Ser Gly Met Leu Met Ala Ala Gly Tyr Gly Trp Ala
 290 295 300
 Gly Leu Gly
 25 305

<210> 36

<211> 927

<212> DNA

30 <213> Synechocystis sp

<400> 36

atggcaacta tccaagcttt ttggcgcttc tcccgccccc ataccatcat tggtaacaact 60
 ctgagcgtct gggctgtgta tctgttaact attctcgggg atggaaactc agttaactcc 120
 35 cctgcttccc tggatttagt gtccggcgct tggctggcct gcctgttggg taatgtgtac 180
 attgtcggcc tcaaccaatt gtgggatgtg gacattgacc gcatcaataa gccgaatttg 240
 ccctagcta acggagattt ttctatcgcc cagggccggt ggattgtggg actttgtggc 300
 gttgcttcct tggcgatcgc ctggggatta gggctatggc tggggctaac ggtgggcatt 360
 agtttgatta ttggcacggc ctattcgggtg ccgccagtga ggttaaagcg cttttccctg 420
 40 ctggcggccc tgtgtattct gacggtgcgg ggaattgtgg ttaacttggg cttattttta 480
 ttttttagaa ttggtttagg ttatcccccc actttaataa ccccatctg ggttttgact 540
 ttatttatct tagttttcac cgtggcgatc gccattttta aagatgtgcc agatatggaa 600
 ggcgatcggc aatttaagat tcaaacttta actttgcaaa tcggcaaaca aaacgttttt 660
 cggggaacct taattttact cactggttgt tatttagcca tggcaatctg gggcttatgg 720
 45 gcggctatgc ctttaaatac tgctttcttg attgtttccc atttgtgctt attagcctta 780
 ctctgggtggc ggagtcgaga tgtacactta gaaagcaaaa ccgaaattgc tagtttttat 840

cagttttattt ggaagctatt tttcttagag tacttgctgt atcccttggc tctgtggta 900
 cctaattttt ctaatactat tttttag 927

<210> 37

5 <211> 308

<212> PRT

<213> Synechocystis sp

<400> 37

10 Met Ala Thr Ile Gln Ala Phe Trp Arg Phe Ser Arg Pro His Thr Ile
 1 5 10 15
 Ile Gly Thr Thr Leu Ser Val Trp Ala Val Tyr Leu Leu Thr Ile Leu
 20 25 30
 Gly Asp Gly Asn Ser Val Asn Ser Pro Ala Ser Leu Asp Leu Val Phe
 15 35 40 45
 Gly Ala Trp Leu Ala Cys Leu Leu Gly Asn Val Tyr Ile Val Gly Leu
 50 55 60
 Asn Gln Leu Trp Asp Val Asp Ile Asp Arg Ile Asn Lys Pro Asn Leu
 65 70 75 80
 20 Pro Leu Ala Asn Gly Asp Phe Ser Ile Ala Gln Gly Arg Trp Ile Val
 85 90 95
 Gly Leu Cys Gly Val Ala Ser Leu Ala Ile Ala Trp Gly Leu Gly Leu
 100 105 110
 Trp Leu Gly Leu Thr Val Gly Ile Ser Leu Ile Ile Gly Thr Ala Tyr
 25 115 120 125
 Ser Val Pro Pro Val Arg Leu Lys Arg Phe Ser Leu Leu Ala Ala Leu
 130 135 140
 Cys Ile Leu Thr Val Arg Gly Ile Val Val Asn Leu Gly Leu Phe Leu
 145 150 155 160
 30 Phe Phe Arg Ile Gly Leu Gly Tyr Pro Pro Thr Leu Ile Thr Pro Ile
 165 170 175
 Trp Val Leu Thr Leu Phe Ile Leu Val Phe Thr Val Ala Ile Ala Ile
 180 185 190
 Phe Lys Asp Val Pro Asp Met Glu Gly Asp Arg Gln Phe Lys Ile Gln
 35 195 200 205
 Thr Leu Thr Leu Gln Ile Gly Lys Gln Asn Val Phe Arg Gly Thr Leu
 210 215 220
 Ile Leu Leu Thr Gly Cys Tyr Leu Ala Met Ala Ile Trp Gly Leu Trp
 225 230 235 240
 40 Ala Ala Met Pro Leu Asn Thr Ala Phe Leu Ile Val Ser His Leu Cys
 245 250 255
 Leu Leu Ala Leu Leu Trp Trp Arg Ser Arg Asp Val His Leu Glu Ser
 260 265 270
 Lys Thr Glu Ile Ala Ser Phe Tyr Gln Phe Ile Trp Lys Leu Phe Phe
 45 275 280 285
 Leu Glu Tyr Leu Leu Tyr Pro Leu Ala Leu Trp Leu Pro Asn Phe Ser

290 Asn Thr Ile Phe
305

5 <210> 38
<211> 1092
<212> DNA
<213> *Synechocystis* sp

10 <400> 38
atgaaatttc cgccccacag tggttaccat tggcaaggct aatcaccttt ctttgaagggt 60
tgggtacgtgc gcttgccttt gccccaatcc ggggaaagtt ttgcttttat gtactccatc 120
gaaaatcctg ctagcgatca tcattacggc ggcgggtgctg tgcaaatttt agggccggct 180
acgaaaaaac aagaaaatca ggaagacca cttgtttggc ggacatttcc ctcggtaaaa 240
15 aaattttggg ccagtcctcg ccagtttgcc ctagggcatt ggggaaaatg tagggataac 300
aggcaggcga aacccctact ctccgaagaa ttttttgcca cgggtcaagga aggttatcaa 360
atccatcaaa atcagcacca aggacaaatc attcatggcg atcgccattg tcgttggcag 420
ttcacgtag aaccggaagt aacttggggg agtcctaacc gatttcctcg ggctacagcg 480
ggttggcttt cctttttacc cttgtttgat cccggttggc aaattctttt agcccaagggt 540
20 agagcgcacg gctggctgaa atggcagagg gaacagtatg aatttgacca cgccctagtt 600
tatgccgaaa aaaattgggg tcaactcctt ccctcccgct ggttttggct ccaagcaaat 660
tattttcctg accatccagg actgagcgtc actgccgctg gcggggaacg gattgttctt 720
ggtcgccccg aagaggtagc tttaattggc ttacatcacc aaggtaattt ttacgaattt 780
ggcccggggc atggcacagt cacttggtgaa gtagctccct ggggcccgtt gcaattaaaa 840
25 gccagcaatg ataggtattg ggtcaagttg tccggaaaaa cagataaaaa aggcagttta 900
gtccacactc ccaccgcca gggcttacaa ctcaactgcc gagataccac taggggctat 960
ttgtatttgc aattgggatc tgtgggtcac ggcctgatag tgcaagggga aacggacacc 1020
gcggggctag aagttggagg tgattggggg ttaacagagg aaaatttgag caaaaaaaca 1080
gtgccattct ga 1092

30 <210> 39
<211> 363
<212> PRT
<213> *Synechocystis* sp

35 <400> 39
Met Lys Phe Pro Pro His Ser Gly Tyr His Trp Gln Gly Gln Ser Pro
1 5 10 15
Phe Phe Glu Gly Trp Tyr Val Arg Leu Leu Leu Pro Gln Ser Gly Glu
40 20 25 30
Ser Phe Ala Phe Met Tyr Ser Ile Glu Asn Pro Ala Ser Asp His His
35 40 45
Tyr Gly Gly Gly Ala Val Gln Ile Leu Gly Pro Ala Thr Lys Lys Gln
50 55 60
45 Glu Asn Gln Glu Asp Gln Leu Val Trp Arg Thr Phe Pro Ser Val Lys
65 70 75 80

	Lys	Phe	Trp	Ala	Ser	Pro	Arg	Gln	Phe	Ala	Leu	Gly	His	Trp	Gly	Lys	
					85					90						95	
	Cys	Arg	Asp	Asn	Arg	Gln	Ala	Lys	Pro	Leu	Leu	Ser	Glu	Glu	Phe	Phe	
				100					105					110			
5	Ala	Thr	Val	Lys	Glu	Gly	Tyr	Gln	Ile	His	Gln	Asn	Gln	His	Gln	Gly	
				115				120					125				
	Gln	Ile	Ile	His	Gly	Asp	Arg	His	Cys	Arg	Trp	Gln	Phe	Thr	Val	Glu	
				130			135					140					
	Pro	Glu	Val	Thr	Trp	Gly	Ser	Pro	Asn	Arg	Phe	Pro	Arg	Ala	Thr	Ala	
10						150					155					160	
	Gly	Trp	Leu	Ser	Phe	Leu	Pro	Leu	Phe	Asp	Pro	Gly	Trp	Gln	Ile	Leu	
					165					170					175		
	Leu	Ala	Gln	Gly	Arg	Ala	His	Gly	Trp	Leu	Lys	Trp	Gln	Arg	Glu	Gln	
				180				185						190			
15	Tyr	Glu	Phe	Asp	His	Ala	Leu	Val	Tyr	Ala	Glu	Lys	Asn	Trp	Gly	His	
				195				200					205				
	Ser	Phe	Pro	Ser	Arg	Trp	Phe	Trp	Leu	Gln	Ala	Asn	Tyr	Phe	Pro	Asp	
				210			215					220					
	His	Pro	Gly	Leu	Ser	Val	Thr	Ala	Ala	Gly	Gly	Glu	Arg	Ile	Val	Leu	
20						230					235					240	
	Gly	Arg	Pro	Glu	Glu	Val	Ala	Leu	Ile	Gly	Leu	His	His	Gln	Gly	Asn	
					245					250					255		
	Phe	Tyr	Glu	Phe	Gly	Pro	Gly	His	Gly	Thr	Val	Thr	Trp	Gln	Val	Ala	
				260				265						270			
25	Pro	Trp	Gly	Arg	Trp	Gln	Leu	Lys	Ala	Ser	Asn	Asp	Arg	Tyr	Trp	Val	
				275				280					285				
	Lys	Leu	Ser	Gly	Lys	Thr	Asp	Lys	Lys	Gly	Ser	Leu	Val	His	Thr	Pro	
				290			295					300					
	Thr	Ala	Gln	Gly	Leu	Gln	Leu	Asn	Cys	Arg	Asp	Thr	Thr	Arg	Gly	Tyr	
30						310					315					320	
	Leu	Tyr	Leu	Gln	Leu	Gly	Ser	Val	Gly	His	Gly	Leu	Ile	Val	Gln	Gly	
					325					330					335		
	Glu	Thr	Asp	Thr	Ala	Gly	Leu	Glu	Val	Gly	Gly	Asp	Trp	Gly	Leu	Thr	
				340				345						350			
35	Glu	Glu	Asn	Leu	Ser	Lys	Lys	Thr	Val	Pro	Phe						
				355				360									

```

    <210> 40
    <211> 56
40  <212> DNA
    <213> Artificial Sequence

```

45

```
<400> 40
cgcgatttaa atggcgcgcc ctgcaggcgg cgcctgcag ggcgcgcat ttaat
<210> 41
```

56

<211> 32
<212> DNA
<213> Artifical Sequence

5 <400> 41
tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 42
<211> 32
10 <212> DNA
<213> Artifical Sequence

<400> 42
tcgacctgca ggaagcttgc ggccgcggat cc 32

15 <210> 43
<211> 32
<212> DNA
<213> Artifical Sequence

20 <400> 43
tcgacctgca ggaagcttgc ggccgcggat cc 32

<210> 44
25 <211> 32
<212> DNA
<213> Artifical Sequence

<400> 44
30 tcgaggatcc gcggccgcaa gcttcctgca gg 32

<210> 45
<211> 36
<212> DNA
35 <213> Artifical Sequence

<400> 45
tcgaggatcc gcggccgcaa gcttcctgca ggagct 36

40 <210> 46
<211> 28
<212> DNA
<213> Artifical Sequence

45 <400> 46
cctgcaggaa gcttgcggcc gcggatcc 28

<210> 47
<211> 36
<212> DNA
5 <213> Artificial Sequence

<400> 47
tcgacctgca ggaagcttgc ggccgcggat ccagct 36

10 <210> 48
<211> 28
<212> DNA
<213> Artificial Sequence

15 <400> 48
ggatccgcgg ccgcaagctt cctgcagg 28

<210> 49
<211> 39
20 <212> DNA
<213> Artificial Sequence

<400> 49
gatcacctgc aggaagcttg cggccgcgga tccaatgca 39

25 <210> 50
<211> 31
<212> DNA
<213> Artificial Sequence

30 <400> 50
ttggatccgc ggccgcaagc ttcctgcagg t 31

<210> 51
35 <211> 41
<212> DNA
<213> Artificial Sequence

<400> 51
40 ggatccgcgg ccgcacaatg gagtctctgc tctctagttc t 41

<210> 52
<211> 38
<212> DNA
45 <213> Artificial Sequence

<400> 52
 ggatcctgca ggtcacttca aaaaaggtaa cagcaagt 38

<210> 53
 5 <211> 45
 <212> DNA
 <213> Artifical Sequence

<400> 53
 10 ggatccgcgg ccgcacaatg gcgttttttg ggctctcccg tgttt 45

<210> 54
 <211> 40
 <212> DNA
 15 <213> Artifical Sequence

<400> 54
 ggatcctgca gggtattgaa aactttcttc aagtacaact 40

20 <210> 55
 <211> 38
 <212> DNA
 <213> Artifical Sequence

25 <400> 55
 ggatccgcgg ccgcacaatg tggcgaagat ctgttggt 38

<210> 56
 <211> 37
 30 <212> DNA
 <213> Artifical Sequence

<400> 56
 35 ggatcctgca ggtcatggag agtagaagga aggagct 37

<210> 57
 <211> 50
 <212> DNA
 <213> Artifical Sequence

40 <400> 57
 ggatccgcgg ccgcacaatg gtacttgccg aggttccaaa gcttgctct 50

<210> 58
 45 <211> 38
 <212> DNA

<213> Artificial Sequence

 <400> 58
 ggatcctgca ggtcacttgt ttctggtgat gactctat 38
 5
 <210> 59
 <211> 38
 <212> DNA
 <213> Artificial Sequence
 10
 <400> 59
 ggatccgagg ccgcacaatg atttcgattc tcaacact 38

 <210> 60
 15 <211> 36
 <212> DNA
 <213> Artificial Sequence

 <400> 60
 20 ggatcctgca ggtcagtgtt gcgatgctaa tgccgt 36

 <210> 61
 <211> 22
 <212> DNA
 25 <213> Artificial Sequence

 <400> 61
 taatgtgtac attgtcggcc tc 22

 30 <210> 62
 <211> 60
 <212> DNA
 <213> Artificial Sequence

 35 <400> 62
 gcaatgtaac atcagagatt ttgagacaca acgtggcttt ccacaattcc ccgcaccgtc 60

 <210> 63
 <211> 22
 40 <212> DNA
 <213> Artificial Sequence

 <400> 63
 aggctaataa gcacaaatgg ga 22
 45
 <210> 64

<211> 63
 <212> DNA
 <213> Artificial Sequence

5 <400> 64
 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggaattgg tttagggttat 60
 ccc 63

<210> 65
 10 <211> 26
 <212> DNA
 <213> Artificial Sequence

<400> 65
 15 ggatccatgg ttgcccaaac cccatc 26

<210> 66
 <211> 61
 <212> DNA
 20 <213> Artificial Sequence

<400> 66
 gcaatgtaac atcagagatt ttgagacaca acgtggcttt gggtaagcaa caatgaccgg 60
 c 61

25 <210> 67
 <211> 25
 <212> DNA
 <213> Artificial Sequence

30 <400> 67
 gaattctcaa agccagccca gtaac 25

<210> 68
 35 <211> 63
 <212> DNA
 <213> Artificial Sequence

<400> 68
 40 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcgggtgcga aaagggtttt 60
 ccc 63

<210> 69
 <211> 23
 45 <212> DNA
 <213> Artificial Sequence

<400> 69
 ccagtgggttt aggctgtgtg gtc 23

5 <210> 70
 <211> 21
 <212> DNA
 <213> Artifical Sequence

10 <400> 70
 ctgagttgga tgtattggat c 21

<210> 71
 <211> 28
 15 <212> DNA
 <213> Artifical Sequence

<400> 71
 ggatccatgg ttacttcgac aaaaatcc 28

20
 <210> 72
 <211> 60
 <212> DNA
 <213> Artifical Sequence

25
 <400> 72
 gcaatgtaac atcagagatt ttgagacaca acgtggcttt gctaggcaac cgcttagtac 60

<210> 73
 30 <211> 28
 <212> DNA
 <213> Artifical Sequence

<400> 73
 35 gaattcttaa cccaacagta aagttccc 28

<210> 74
 <211> 63
 <212> DNA
 40 <213> Artifical Sequence

<400> 74
 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcgccggcat tgtcttttac 60
 atg 63

45
 <210> 75

<211> 20
 <212> DNA
 <213> Artificial Sequence

5 <400> 75
 ggaacccttg cagccgcttc 20

<210> 76
 <211> 22
 10 <212> DNA
 <213> Artificial Sequence

<400> 76
 gtatgcccaa ctggtgcaga gg 22

15 <210> 77
 <211> 28
 <212> DNA
 <213> Artificial Sequence

20 <400> 77
 ggatccatgt ctgacacaca aaataccg 28

<210> 78
 25 <211> 62
 <212> DNA
 <213> Artificial Sequence

<400> 78
 30 gcaatgtaac atcagagatt ttgagacaca acgtggcttt cgccaatacc agccaccaac 60
 ag 62

<210> 79
 <211> 27
 35 <212> DNA
 <213> Artificial Sequence

<400> 79
 gaattctcaa atccccgcat ggcctag 27

40 <210> 80
 <211> 65
 <212> DNA
 <213> Artificial Sequence

45 <400> 80

ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggcctacg gcttggacgt 60
 gtggg 65

<210> 81
 5 <211> 21
 <212> DNA
 <213> Artifical Sequence

<400> 81
 10 cacttggatt cccctgatct g 21

<210> 82
 <211> 21
 <212> DNA
 15 <213> Artifical Sequence

<400> 82
 gcaataccg cttggaaaac g 21

20 <210> 83
 <211> 29
 <212> DNA
 <213> Artifical Sequence

25 <400> 83
 ggatccatga ccgaatcttc gccctagc 29

<210> 84
 <211> 61
 30 <212> DNA
 <213> Artifical Sequence

<400> 84
 gcaatgtaac atcagagatt ttgagacaca acgtggcttt caatcctagg tagccgaggc 60
 35 g 61

<210> 85
 <211> 27
 <212> DNA
 40 <213> Artifical Sequence

<400> 85
 gaattcttag cccaggccag cccagcc 27

45 <210> 86
 <211> 66

<212> DNA
 <213> Artifical Sequence

<400> 86
 5 ggtatgagtc agcaacacct tcttcacgag gcagacctca gcggggaatt gatttggtta 60
 attacc 66

<210> 87
 <211> 21
 10 <212> DNA
 <213> Artifical Sequence

<400> 87
 15 gcgatcgcca ttatcgcttg g 21

<210> 88
 <211> 24
 <212> DNA
 <213> Artifical Sequence
 20

<400> 88
 gcagactggc aattatcagt aacg 24

<210> 89
 25 <211> 25
 <212> DNA
 <213> Artifical Sequence

<400> 89
 30 ccatggattc gagtaaagtt gtcgc 25

<210> 90
 <211> 0
 <213> Artifical Sequence
 35

<400> 90
 gaattcactt caaaaaaggt aacag

<210> 91
 40 <211> 4550
 <212> DNA
 <213> Arabidopsis sp

<400> 91
 45 attttacacc aatttgatca cttaactaaa ttaattaaat tagatgatta tcccaccata 60
 tttttgagca ttaaaccata aaaccatagt tataagtaac tgttttaatc gaatatgact 120

	cgattaagat	taggaaaaat	ttataaccgg	taattaagaa	aacattaacc	gtagtaaccg	180
	taaatgccga	ttcctccctt	gtctaaaaga	cagaaaacat	atattttatt	ttgccccata	240
	tgtttctactc	tatttaattt	caggcacaat	acttttggtt	ggtaacaaaa	ctaaaaagga	300
	caacacgtga	tacttttcct	cgtccgtcag	tcagattttt	tttaaactag	aaacaagtgg	360
5	caaactctaca	ccacattttt	tgtttaatct	attaacttgt	aagttttaaa	ttcctaaaaa	420
	agtctaacta	attcttctaa	tataagtaca	ttccctaaat	ttcccaaaaa	gtcaaattaa	480
	taattttcaa	aatctaactt	aaatatctaa	taattcaaaa	tcattaaaaa	gacacgcaac	540
	aatgacacca	attaatcatc	ctcgaccac	acaattctac	agttctcatg	ctaaaccata	600
	ttttttgtct	tctgttcctt	caaaatcatt	tctttctctt	ctttgattcc	caaagatcac	660
10	ttctttgtct	ttgatttttg	atTTTTTTTt	tctctggcgt	gaaggaagaa	gctttatttc	720
	atggagtctc	tgctctctag	ttcttctctt	gtttccgctg	gtaaatctcg	tccttttctg	780
	gtttcaggtt	ttatttggtg	tttaggtttc	gtttttgtga	ttcagaacca	tacaaaaagt	840
	ttgaactttt	ctgaatataa	aataaggaaa	aagtttcgat	ttttataatg	aattgtttac	900
	tagatcgaag	taggtgacaa	aggttattgt	gtggagaagc	ataatttctg	ggcttgactt	960
15	tgaattttgt	ttctcatgca	tgcaacttat	caatcagctg	gtgggttttg	ttggaagaag	1020
	cagaatctaa	agctccactc	tttatcaggt	tcgttagggg	tttatgggtt	tttgaaatta	1080
	aatactcaat	catcttagtc	tcattattct	attgggtgaa	tcacattttc	taatttgga	1140
	tttatgagac	aatgtatgtt	ggacttaggt	gaagttcttc	tctttgggta	tagttgaagt	1200
	gttactgatg	ttgttttagct	ctttacacca	atatatacac	ccaattttgc	agaaatccga	1260
20	gttctgcgtt	gtgattcgag	taaagtgtgc	gcaaaaccga	agtttaggaa	caatcttggt	1320
	aggcctgatg	gtcaaggatc	ttcattgttg	ttgtatccaa	aacataagtc	gagatttcgg	1380
	gttaatgcc	ctgcgggtca	gcctgaggct	ttcgactcga	atagcaaaca	gaagtctttt	1440
	agagactcgt	tagatgcgtt	ttacagggtt	tctaggcctc	atacagttat	tggcacagtt	1500
	aagtttctct	ttaaaaatgt	aactctttta	aaacgcaatc	tttcagggtt	ttcaaggaga	1560
25	taacattagc	tctgtgattg	gatttgcagg	tgcttagcat	tttatctgta	tctttcttag	1620
	cagtagagaa	ggtttctgat	atatctcctt	tacttttcac	tggcatcttg	gaggtaatga	1680
	atatataaca	cataatgacc	gatgaagaag	atacattttt	ttcgtctctc	tgtttaaaca	1740
	attgggtttt	gttttcaggc	tgttgttgca	gctctcatga	tgaacattta	catagttggg	1800
	ctaaatcagt	tgtctgatgt	tgaaatagat	aaggtaacat	gcaaattttc	ttcatatgag	1860
30	ttcgagagac	tgatgagatt	aatagcagct	agtgcctaga	tcatctctat	gtgggttttt	1920
	gcagggttaac	aagccctatc	ttccattggc	atcaggagaa	tattctgtta	acaccggcat	1980
	tgcaatagta	gcttccttct	ccatcatggt	atgggtgccat	tttcacaaaa	tttcaacttt	2040
	tagaattcta	taagttactg	aaatagtttg	ttataaatcg	ttatagagtt	tctggcttgg	2100
	gtggattggt	ggttcatggc	cattgttctg	ggctcttttt	gtgagtttca	tgctcggtag	2160
35	tgcatactct	atcaatgtaa	gtaagtttct	caatactaga	atttggctca	aatcaaaatc	2220
	tgcagtttct	agtttttaggt	taatgagggt	ttaataactt	acttctacta	caaacagttg	2280
	ccactttttac	ggtggaaaag	atttgcattg	gttgagcga	tgtgtatcct	cgctgtccga	2340
	gctattattg	ttcaaatcgc	cttttatctt	catattcagg	tactaaacca	ttttccttat	2400
	gtttttgtagt	tgttttcatc	aaaatcactt	ttatattact	aaagctgtga	aactttgttg	2460
40	cagacacatg	tgtttggaag	accaatcttg	ttcactaggc	ctcttatttt	cgccactgcg	2520
	tttatgagct	ttttctctgt	cgttattgca	ttgtttaagg	taaacaaaga	tggaaaaaga	2580
	ttaaatctat	gtatacttaa	agtaaagcat	tctactgtta	ttgatgagaa	gttttctttt	2640
	ttggttggat	gcaggatata	cctgatatcg	aaggggataa	gatattcgga	atccgatcat	2700
	tctctgtaac	tctgggtcag	aaacgggtac	gatattctaa	ctaaagaaat	tgttttgact	2760
45	caagtgttgg	attaagatta	cagaagaaag	aaaactgttt	ttgtttcttg	caaaattcag	2820
	gtgttttgg	catgtgttac	actacttcaa	atgggttacg	ctgttgcaat	tctagttgga	2880

	gccacatctc cattcatatg gagcaaagtc atctcggtaa caatctttct ttacccatcg	2940
	aaaactcgct aattcatcgt ttgagtggtta ctggtttcat tttgttccgt tctgttgatt	3000
	ttttttcagg ttgtgggtca tgttatactc gcaacaactt tgtgggctcg agctaagtcc	3060
	gttgatctga gtagcaaaac cgaaataact tcatgttata tgttcatatg gaaggtaga	3120
5	ttcgtttata aatagagtct ttactgcctt tttatgcgct ccaatttggga attaaaatag	3180
	ccttttcagtt tcacgaatc accattatac tgataaattc tcatttctgc atcagctctt	3240
	ttatgcagag tacttgctgt tacctttttt gaagtgactg acattagaag agaagaagat	3300
	ggagataaaa gaataagtca tcactatgct tctgttttta ttacaagttc atgaaattag	3360
	gtagtgaact agtgaattag agttttattc tgaaacatgg cagactgcaa aaatatgtca	3420
10	aagatatgaa tttctgttgg gtaaagaagt ctctgcttgg gcaaaatctt aagggttcggt	3480
	gtgttgatat aatgctaagc gaagaaatcg attctatgta gaaatttccg aaactatgtg	3540
	taaacatgtc agaacatctc cattctatat ctcttctgc aagaaagctc tgtttttatc	3600
	acctaaactc tttatctctg ttagttaaag atatgtatat gtacgtgact acattttttt	3660
	gttgatgtaa tttgcagaac gtatggattt ttgttagaaa gcatgagttc gaaagtatat	3720
15	gtttatatat atggataatt cagacctaac gtcgaagctc acaagcataa attcactact	3780
	atagtttgct ctgtaataga tagttccatt gatgtcttga aactgtacgt aactgcctgg	3840
	gcgttttgtg gttgatactg actactgagt gttctttgtg agtggtgtaa gtatacaaga	3900
	agaagaatat aggtcacgg gaacgactgt ggtggaagat gaaatggaga tcatcacgta	3960
	gcggctttgc caaagaccga gtcacgatcg agtctatgaa gtctttacag ctgctgatta	4020
20	tgattgacca ttgcttagag acgcattgga atcttactag ggacttgcct gggagtttct	4080
	tcaagtacgt gtcagatcat acgatgtagg agatttcacg gctttgatgt gtttgtttgg	4140
	agtcacaatg cttaatgggc ttattggccc aataatagct agctcttttg ctttagccgt	4200
	ttcgtttgtc ccctggtggt gagtattatt agggatggt gtgaccaaag tcaccagacc	4260
	tagagtgaat ctagtagagt cctagaccat ggtccatggc ttttatttgt aatttgaaaa	4320
25	atgaacaatt ctttttgtaa ggaaaacttt tatatagtag acgtttacta tatagaaact	4380
	agttgaacta acttcgtgca attgcataat aatgggtgta aatagagggt gcaaaactca	4440
	ataaacattt cgacgtacca agagttcgaa acaataagca aaatagattt ttttgcttca	4500
	gactaatttg tacaatgaat ggttaataaa ccattgaagc ttttattaat	4550
30	<210> 92	
	<211> 4450	
	<212> DNA	
	<213> Arabidopsis sp	
35	<400> 92	
	tttaggttac aaaatcaatg atattgcgta tgtcaactat aaaagccaaa agtaaagcct	60
	cttgtttgac cagaaggcca tgatcattgt atacatacag ccaaactacc tcctggaaga	120
	aaagacatgg atcccaaaca acaacaatag cttcttttac aagaaccagt agtaactagt	180
	cactaatcta aaagagttaa gtttcagctt ttctggcaat ggctccttga tcatttcaat	240
40	cctgaaggag acccactttg tagcaagacc atgtcctctg ttccacttac agtgtgtctc	300
	aaaagtctac ttcaattctt catatatagg ttccctcacac tacagcttca tcctcattcg	360
	ttgacagaga gagagtcttt attgaaaact tcttccaagt acaactccac taaatataat	420
	agcaccaaac cacttgttcg acacaaatct gtacagatat aaaaacacta ttaggttttc	480
	caaggcaaat cacataattg gattgtgaaa gagtacaaaa gataaaacca aattttcata	540
45	ctttctactg cagtcagcac cagatgataa gtcagctgtc cctatttggc atcctaactg	600
	tcctgatgca gcggccagtg atgcgtaata ttgccaccct taatcattag agcgagaac	660

	aaaaagaatc	aaaagacagt	aatggaatt	aggaatcaca	aatgagtcct	tgtaaagttt	720
	attgagtagc	gagatctgca	ctgaatccag	aaagtgcag	aaaacctatg	gatgctgtgc	780
	caaatccagt	taaccaaagc	tttgtattat	caccgaatct	aagggctggt	gacttaacac	840
	caactttttac	atcatcttct	ttgtcctgga	gacacaatat	attagacatt	agtccatgga	900
5	aaaaaatga	tttaacctag	aatatctcaa	aattacttgc	ataaaaactg	aacttgagct	960
	gaaattttgg	gttcgtagct	tgtggcatat	actatttcat	tttcaatggg	ccacaaaggt	1020
	aactttcttt	tctcacttct	gttgcaaagc	ggaagacttt	tatggggcta	actcttcact	1080
	taaagtatat	aaatcagatg	gaaaagggtg	gagatcaggg	taattttctt	ctttatgatt	1140
	gacaaaagtc	gaacatcgaa	atggatgcat	ttgcatgaga	catgaaacaa	aagctgaaaa	1200
10	agaaatctgt	gggtgtgaag	ctagaaaaag	aaaacaaagc	aagcaatatg	cacacattga	1260
	gattaactac	tttgctactg	gtcataatca	aatagatttt	gaagctaaaa	aataaaaagt	1320
	gaatatacct	gatgtgcata	aatagtatca	taaacaaggg	tccagcagac	tccggagaga	1380
	tagagaggga	gtacaataga	tggtgctatg	cttcctttta	ctgcagtcca	tcctaacaat	1440
	gctccccagt	ttatgggtcaa	acctaanaag	gcttgaggct	gcaattataa	aaacgaatca	1500
15	atcataagaa	aatcagaaaa	tatataatgt	ctaactttga	gaagccagaa	tagattttaa	1560
	ttacccaaaa	tgtaaacctc	ttcataagtg	ggtaggaaaa	gacaagtaac	aaagatgaag	1620
	cccctaaaa	acggctgcag	aatatacata	ctgaaatgag	ctcaagtaga	aaagaatttg	1680
	atcacaaaa	taaagacaag	acctgagaac	atatcttcag	aatttggggc	aactacataa	1740
	gggtgaacca	tatgtgtatg	tgaattttta	aacaaacact	tgcaaatacg	cgacttttag	1800
20	gcaagtaaaa	aatccaaaca	aacctgtaat	tgttaagttg	gagaagaatc	cctaagccta	1860
	aaagcaactg	cagcccgaga	aatccaatcc	cttgaaatgg	tgtcaaaaga	ccactggcga	1920
	taggtcttag	ttttgtacga	tcaacctgga	tataaaagaa	atttgtaaga	caacataatc	1980
	taaaacaaaa	caaccataca	aaatcttgag	ctttacatac	aagcaacca	tctttgttta	2040
	tggaagaatg	aatccagtta	catgaatgct	gtgtatctac	cctaactact	aaacacatat	2100
25	ttcaatcgaa	aaacatatcc	caccttcacc	atatctaaca	cctgaagtct	ttcacttttt	2160
	gaacgaagtc	atcagaacat	gcagataagc	tattacccaa	aacagagata	tgactggaaa	2220
	tggtgtcgta	aattgatcca	acatagaaaa	atcaagacca	gttccagatg	tcaaagcaat	2280
	aacactttcc	caccatgggt	acagaaacca	tagttacaca	aaacatgttt	cctaaaccaa	2340
	cataactaa	ggatatataa	atttgacatc	actttatcac	cataccataa	gatagcttaa	2400
30	aaacaaactg	acctttgtat	ctatgtcctg	atcaagcaga	tcattttatag	tacaaccagc	2460
	acctctaaga	agtaatgctc	cgcaacccaa	taaagccata	tattttaa	ttggaaggct	2520
	tccaggatca	gcagccaacg	caatcgacct	atacaacaat	gatggagatt	cagagtatcg	2580
	atctatttac	atagctctgg	aactagatcc	atgacgaaac	atggaacatc	gttataatat	2640
	ctaaagactt	ccaaacagat	tcctgagtaa	gaaacccagt	ggaactatag	tactgtaaca	2700
35	tatataaaat	caaagaaaac	tcaggtttat	agcattatcc	aatcctgatt	tctgccaatc	2760
	cttaaccact	ctcccatgct	atcaaaaacc	tcagctcaag	atcatactac	ctaattgcct	2820
	atgagctctt	gggaagatca	ttatggattt	gataactgaa	aaaagtaaca	gagaaatagc	2880
	agactgcaag	aactactcca	aacttctcca	ctgatatgta	tgtagtctaa	caataataaa	2940
	cagacataaa	ttcttttatc	aagcttcaag	agcaagttag	tcagaaaaca	tcacagccaa	3000
40	accaaccagg	aaaacacata	actttatcac	ataaaactaa	atttaagtga	atctgactta	3060
	acataaacca	tcctttggga	cgaaaggaaa	ctatataaac	atgcagtctt	tccttccctc	3120
	agctattctt	tcggatggat	tataatgaat	ctcaaaagtg	aaatgtcttg	attctcagct	3180
	acattactca	aaggcgaaga	taaacttacc	acatacaagg	ccacgcaagc	aaccaagttc	3240
	caatgggttt	atccaatcga	gcaagcttag	cataacctct	aacttcttct	ggtaaatata	3300
45	aatctatcca	agaagcttcc	ttaacaacaa	caccatcact	cttctcctta	tcacttttct	3360
	tcggctttcc	ctccaaaacc	gaagaagacg	acgacattcc	acaaattaat	ctgtaattcc	3420

	aaccaacacc	aaaaaacttc	tccatgatgca	attctcttcc	tttactccat	acttggtaat	3480
	tatcattcca	tgaaggataa	cacttagtga	aaggatttgt	gtaatgggta	gtcacaggat	3540
	tggaacaagga	tttatgttgt	gattgcaaaa	gagcagagga	agaagatgga	gttacggaga	3600
	cggaagattt	caacaaccgt	cttgaaacac	gggagagccc	aaaaaacgcc	atctttgaga	3660
5	gaaattgttg	cctggaagaa	acaaagactt	gagatttcaa	acgtaagtga	attcttacga	3720
	acgaaagcta	acttctcaag	agaatcagat	tagtgattcc	tcaaaaacaa	acaaaactat	3780
	ctaatttcag	tttcgagtga	tgaagcctta	agaatctaga	acctccatgg	cgtttcta	3840
	ctctcagaga	taatcgaatt	ccttaaacaa	tcaaagctta	gaaagagaag	aacaacaaca	3900
	acaacaaaaa	aaatcagatt	aacaaccgac	cagagagcaa	cgacgacgcc	ggcgagaaag	3960
10	agcacgtcgt	ctcggagcaa	gacttcttct	ccagtaaccc	ggatggatcg	ttaatgggcc	4020
	tgtagattat	tatatgttgg	ccgaaacaat	tgggtcagca	aaaacttggg	ggataatgaa	4080
	gaaacacgta	cagtatgcat	ttaggtccca	aattaattgg	ccatataatt	cgaatcagat	4140
	aaactaatca	acctctacct	tacttatttc	tcactgtttt	tatttctacc	ttagtagttg	4200
	aagaaacact	tttatttatc	ttttcgggac	ccaaatttga	taggatcggg	ccattactca	4260
15	tgagcgtcag	acacatatta	gccttatcag	attagtgggg	taagggtttt	ttaattcggt	4320
	aagaagcaac	aatcaatgtc	ggagaaatta	aagaatctgc	atgggcgtgg	cgatgatgata	4380
	tgtgcatatg	gagtcagttg	ccgatcatat	ataactattt	ataaactaca	tataaagact	4440
	actaatagat						4450
20	<210>	93					
	<211>	2850					
	<212>	DNA					
	<213>	Arabidopsis sp					
25	<400>	93					
	aattaaaatt	tgagcgggtct	aaaccattag	accgtttaga	gatccctcca	acccaaaata	60
	gtcgattttc	acgtcttgaa	catatattgg	gccttaattct	gtgtgggttag	taaagacttt	120
	tatttggtcaa	agaaaaacaa	ccatggccca	acatgttgat	acttttat	aattatacaa	180
	gtacccttga	attctctgaa	atataattga	ttgaccaga	tattaatttt	aattatcatt	240
30	tcctgtaaaa	gtgaaggagt	caccgtgact	cgctgtaatc	tgaaccaat	ctgttcatat	300
	gatgaagaag	tttctctcgt	tctcctccaa	cgcgtagaaa	attctgacgg	cttaacgatg	360
	tggcgaagat	ctgttggtta	tgcgttctct	tcaagaatct	ctgtttcttc	ttcgttacca	420
	aaccctagac	tgattccttg	gtcccgcgaa	ttatgtgccg	ttaatagctt	ctccagcct	480
	ccggtctcga	cggaatcaac	tgctaagtta	gggatcactg	gtgttagatc	tgatgccaat	540
35	cgagtttttg	ccactgctac	tgccgccgct	acagctacag	ctaccacggg	tgagatttcg	600
	tctagagtgt	cggctttggc	tggattaggg	catcactacg	ctcgttggtta	ttgggagctt	660
	tctaaagcta	aacttaggta	tgtgtttact	tttcttttct	catgaaaaat	ctgaaaattt	720
	ccaattgttg	gattctttaa	ttctcatttg	ttttatgggt	gtagtatgct	tgtggttgca	780
	acttctggaa	ctgggtatat	tctgggtacg	ggaaatgctg	caattagctt	ccgggggctt	840
40	tgttacacat	gtgcaggaac	catgatgatt	gctgcactcg	ctaattcctt	gaatcaggct	900
	attgaaatgt	tgagaagtgc	ataaatttcg	aatccttggt	gtgtttatgt	agttgatctt	960
	gcttgcttat	gtttatgtag	ttgaaaagtt	taaaaatttc	taatccttgg	tagttgatct	1020
	cgcttgcttg	ttttttcatt	ttctagattt	ttgagataag	caatgattct	aagatgaaaa	1080
	gaacgatgct	aaggccattg	ccttcaggac	gtattagtg	tccacacgct	gttgcatggg	1140
45	ctactattgc	tggtgcttct	ggtgcttggt	tgttggccag	caagggtgaat	gtttgttttt	1200
	ttatatgtga	tttctttgtt	ttatgaatgg	gtgattgaga	gattatggat	ctaaactttt	1260

	gcttccacga caaggttatt gcagactaat atgttggttg ctggacttgc atctgccaat	1320
	cttgactttt atgcgtttgt ttatactccg ttgaagcaac ttcaccctat caatacatgg	1380
	gttggcgctg ttgttggtgc tatccacccc ttgcttgggt aaatttttgt tccttttctt	1440
	ctttatttta gcagattctg ttttggttga tactgctttt aattcaaaat gtagtcatgg	1500
5	ttcaccaatt ctatgcttat ctattttgtg tgtgtgcagg tgggcggcag cgtctggtca	1560
	gatttcatac aattcgatga ttcttccagc tgctctttac ttttggcaga tacctcattt	1620
	tatggccctt gcacatctct gccgcaatga ttatgcagct ggagggttaag accatatggt	1680
	gtcatatgag attagaatgt ctcccttccat gtagtggtga tcttgaacta gttcaatttc	1740
	gtggaatgat cagagtgtcc tagatagtgt cacagcagtc gacattttag tggctagata	1800
10	atgagtctct tccgtagag ataaacattc gcgaacattg tttccagctt ccgcgacca	1860
	acttctgatt ttgtttcttg gtacctgtt ttcagttaca agatgtgtc actctttgat	1920
	ccgtcaggga agagaatagc agcagtggct ctaaggaaact gcttttacct gatccctctc	1980
	ggtttcatcg cctatgactg tgagtcttgt agattcatct tttttttgta gtttattgac	2040
	tgcattgctg tatctgattt ttgctgttcc ttccaatttt tgtgacaggg ggggttaacct	2100
15	caagttggtt ttgcctcgaa tcaacacttc tcacactagc aatcgctgca acagcatttt	2160
	cattctaccg agaccggacc atgcataaag caaggaaaat gttccatgcc agtcttctct	2220
	tccttcctgt tttcatgtct ggtcttcttc tacaccgtgt ctctaattgat aatcagcaac	2280
	aactcgtaga agaagccgga ttaacaaatt ctgtatctgg tgaagtcaaa actcagaggc	2340
	gaaagaaacg tgtggctcaa cctccggtgg cttatgcctc tgctgcaccg tttcctttcc	2400
20	tcccagctcc ttccttctac tctccatgat aacctttaag caagctattg aatttttgga	2460
	aacagaaatt aaaaaaaaaa tctgaaaagt tcttaagttt aatctttggt taataatgaa	2520
	gtggagaacg catacaagtt tatgtatttt ttctcatctc cacataattg tattttttct	2580
	ctaagtatgt ttcaaatgat acaaaataca tactttatca attatctgat caaattgatg	2640
	aatttttgag ctttgacgtg ttaggtctat ctaataaacg tagtaacgaa tttggttttg	2700
25	gaaatgaaat ccgataaccg atgatggtgt agagttaaac gattaaaccg ggttggttaa	2760
	aggtctcgag tctcgacggc tgcggaaatc ggaaaatcac gattgaggac tttgagctgc	2820
	cacgaagatg gcgatgaggt tgaaatcaat	2850
	<210> 94	
30	<211> 3660	
	<212> DNA	
	<213> Arabidopsis sp	
	<400> 94	
35	tatttgattt ttattgtta aattttatga tttcaccggg tatatatcat cccatattaa	60
	tattagattt attttttggg ctttatttgg gttttcgatt taaactgggc ccattctgct	120
	tcaatgaaac cctaattgggt tttgtttggg ctttggattt aaaccgggccc cattctgctt	180
	caatgaaggt cttttgtcca acaaaactaa catccgacac aactagtatt gccaaagagga	240
	tcgtgccaca tggcagttat tgaatcaaag gccgccaaaa ctgtaacgta gacattactt	300
40	atctccggta acggacaacc actcgtttcc cgaaacagca actcacagac tcacaccact	360
	ccagtctccg gcttaactac caccagagac gattctctct tccgtcgggt ctatgacttc	420
	gattctcaac actgtctcca ccatccactc ttccagagtt acctccgtcg atcgagtcgg	480
	agtcctctct cttcggaatt cggattccgt tgagttcact cgccggcggt ctggtttctc	540
	gacgttgatc tacgaatcac ccggtagtta gcattctggt ggatagattg atgaatgttt	600
45	tcttcgattt tttttttact gatcttggtg tggatctctc gtagggcgga gatttgttgt	660
	gcgtgcggcg gagactgata ctgataaagg tatgattttt tagttgtttt tattttctct	720

	ctcttcaaaa	ttctcttttc	aaacactgtg	gcgtttgaat	ttccgacggc	agttaaattct	780
	cagacacctg	acaaggcacc	agccgggtgg	tcaagcatta	accagcttct	cggatatcaaa	840
	ggagcatctc	aagaaactgt	aattttgttc	atctcctcag	aatcttttaa	attatcatat	900
	ttgtggataa	tgatgtgtta	gttttaggaat	tttcctacta	aaggtaattct	cttttgagga	960
5	caagtcttgt	ttttagctta	gaaatgatgt	gaaaatgttg	tttgtagct	aaaaagagtt	1020
	tgttgttata	ttctgtattc	agaataaatg	gaagattcgt	cttcagctta	caaaaccagt	1080
	cacttggcct	ccactggttt	ggggagtcgt	ctgtgggtgct	gctgcttcag	gtaatcatac	1140
	gaacctcttt	tggatcatgc	aatactgtac	agaaagtttt	ttcattttcc	ttccaattgt	1200
	ttcttctggc	agggaaacttt	cattggaccc	cagaggatgt	tgctaagtcg	attctttgca	1260
10	tgatgatgtc	tggtccttgt	cttactggct	atacacaggt	ctggttttac	acaacaaaaa	1320
	gctgacttgt	tcttattcta	gtgcatttgc	ttggtgctac	aataacctag	acttgtcgat	1380
	ttccagacaa	tcaacgactg	gtatgataga	gatatcgacg	caattaatga	gccatatcgt	1440
	ccaattccat	ctggagcaat	atcagagcca	gaggtaactg	agacagaaca	ttgtgagctt	1500
	ttatctcttt	tgtgattctg	atttctcctt	actccttaaa	atgcagggtta	ttacacaagt	1560
15	ctgggtgcta	ttattgggag	gtccttggat	tgctggaata	ttagatgtgt	gggtaagttg	1620
	gcccttctga	cattaactag	tacagttaaa	gggcacatca	gatttgctaa	aatcttccct	1680
	tatcaggcag	ggcataccac	tcccactgtc	ttctatcttg	ctttgggagg	atcattgcta	1740
	tcttatatat	actctgctcc	acctcttaag	gtaagtttta	ttcctaactt	ccactctcta	1800
	gtgataagac	actccatcca	agttttggag	ttttgaatat	cgatatctga	actgatctca	1860
20	ttgcagctaa	aacaaaatgg	atgggttggg	aattttgcac	ttggagcaag	ctatattagt	1920
	ttgccatggt	aagatatctc	gtgtatcaat	aatatatggc	gttggtctca	tctcattgat	1980
	ttgtttcttg	ctcacttgac	tgataggtgg	gctggccaag	cattgtttgg	cactcttacg	2040
	ccagatgttg	ttgttctaac	actcttgtag	agcatagctg	gggtactctt	ttggcaaac	2100
	ttttatgttg	cttttttctg	tatctgttgt	aatatgctct	tgcttcatgt	tgtacctttg	2160
25	tgataatgca	gttaggaata	gccattgtta	acgacttcaa	aagtgttgaa	ggagatagag	2220
	cattaggact	tcagtctctc	ccagtagctt	ttggcaccga	aactgcaaaa	tgatattgag	2280
	ttgggtgctat	agacattact	cagctttctg	ttgccgggtat	gtactatcca	ctgtttttgt	2340
	gcagctgtgg	cttctatttc	ttttccttga	tcttatcaac	tgatatttca	ccaatggtaa	2400
	agcacaaatt	aatgaagctg	aatcaacaaa	ggcaaaacat	aaaagtacat	tctaataaaa	2460
30	tgagctaatt	aagaggaggc	atctactttt	atgtttcatt	agtgtgattg	atggattttc	2520
	atttcatgct	tctaaaacaa	gtattttcaa	cagtgtcatg	aaataacaga	acttatatct	2580
	tcatttgtac	ttttactagt	ggatgagtta	cacaatcatt	gttatagaac	caaatcaaag	2640
	gtagagatca	tcattagtat	atgtctattt	tggttgagg	atatctatta	gcattctggga	2700
	aaccttatta	tgcgttggcg	ttggttgctt	tgatcattcc	tcagattgtg	ttccaggtaa	2760
35	agacgttaac	agtctcacat	tataattaat	caaattcttg	tcactcgtct	gattgctaca	2820
	ctcgcttcta	taaactgcag	tttaataact	ttctcaagga	ccctgtcaaa	tacgacgtca	2880
	agtaccagg	aagtcaactt	agtacacatg	tttgtgttct	tttgaaatat	ctttgagagg	2940
	tctcttaatc	agaagtgtgt	tgaaacactc	atcttgatta	caggcaagcg	cgcagccatt	3000
	cttgggtgctc	ggaatatattg	taacggcatt	agcatcgcaa	caactgaaaa	ggcgtatttt	3060
40	gatgggggtt	tgtcgaaagc	agagggtgtg	acacatcaaa	tgtgggcaag	tgatggcatc	3120
	aactagttta	aaagattttg	taaaatgtat	gtaccgttat	tactagaaac	aactcctgtt	3180
	gtatcaattt	agcaaaacgg	ctgagaaatt	gtaattgatg	ttaccgtatt	tgcgctccat	3240
	ttttgcattt	cctgctcata	tcgaggattg	gggtttatgt	tagttctgtc	acttctctgc	3300
	tttcagaatg	tttttgtttt	ctgtagtgga	ttttaactat	tttcatcact	ttttgtattg	3360
45	attctaataa	tgtatccaca	taaaaacagt	aatatacaaa	aatgataact	cctcaaactt	3420
	tttataatct	aatctaaca	actagctagt	aaccaacta	acttcataca	attaatttga	3480

gaaactacaa	agactagact	atacatatgt	tatttaacaa	cttgaaactg	tggtattact	3540
acctgatttt	tttctattct	acagccattt	gatatgctgc	aatcttaaca	tatcaagtct	3600
cacgttggtg	gacacaacat	actatcacia	gtaagacacg	aagtaaaacc	aaccggcaac	3660

(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES
PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro



(43) Internationales Veröffentlichungsdatum
18. Januar 2001 (18.01.2001)

PCT

(10) Internationale Veröffentlichungsnummer
WO 01/04330 A1

(51) Internationale Patentklassifikation⁷: C12N 15/82, 15/54, 9/10, 15/31, G01N 33/53, A01H 5/00

(21) Internationales Aktenzeichen: PCT/EP00/05862

(22) Internationales Anmeldedatum:
23. Juni 2000 (23.06.2000)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
199 31 834.4 9. Juli 1999 (09.07.1999) DE

(71) Anmelder (für alle Bestimmungsstaaten mit Ausnahme von US): SUNGENE GMBH & CO. KGAA [DE/DE]; D-06468 Gatersleben (DE).

(72) Erfinder; und

(75) Erfinder/Anmelder (nur für US): HERBERS, Karin [DE/DE]; Am Hange 6, D-06484 Quedlinburg (DE). BADUR, Ralf [DE/DE]; Petersilienstrasse 17, D-38640 Goslar (DE). KUNZE, Irene [DE/DE]; Mühlenweg 11, D-06466 Gatersleben (DE). GEIGER, Michael [DE/DE]; Neuer Weg 15, D-06484 Quedlinburg (DE).

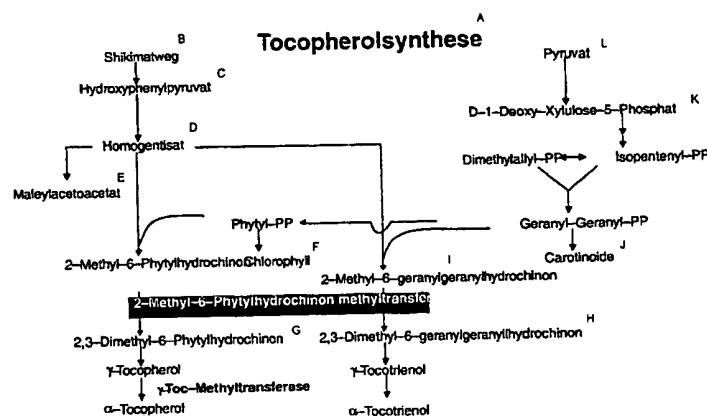
(74) Anwalt: BIEBERBACH, Andreas; BASF Aktiengesellschaft, D-67056 Ludwigshafen (DE).

(81) Bestimmungsstaaten (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

[Fortsetzung auf der nächsten Seite]

(54) Title: IDENTIFICATION AND OVEREXPRESSION OF A DNA SEQUENCE CODING FOR 2-METHYL-6-PHYTYLHYDROQUINONE-METHYLTRANSFERASE IN PLANTS

(54) Bezeichnung: IDENTIFIZIERUNG UND ÜBEREXPRESSION EINER DNA-SEQUENZ KODIEREND FÜR EINE 2-METHYL-6-PHYTYLHYDROCHINON-METHYLTRANSFERASE IN PFLANZEN



WO 01/04330 A1

(57) Abstract: A method for the production of plants with an increased tocopherol and tocotrienol content by overexpression of a gene coding for 2-methyl-6-phytylhydroquinone-methyltransferase.

(57) Zusammenfassung: Verfahren zur Herstellung von Pflanzen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen durch Überexpression eines Gens codierend für eine 2-Methyl-6-phytylhydrochinon-methyltransferase.

REF
A44

Della Penna, et al.
SN 09/560,761

Identifizierung und Überexpression einer DNA-Sequenz codierend für eine 2-Methyl-6-phytylhydrochinon-methyltransferase in Pflanzen

5

Beschreibung

Die Erfindung betrifft eine DNA kodierend für ein Polypeptid mit 2-Methyl-6-phytylhydrochinon-methyltransferase Aktivität. Zudem

10 betrifft die Erfindung die Verwendung von DNA-Sequenzen codierend für ein Polypeptid mit 2-Methyl-6-phytylhydrochinon-Methyltransferase Aktivität zur Herstellung von Pflanzen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen, speziell die Verwendung der DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder mit dieser hybri-

15 disierende oder zur Gesamtsequenz oder zu Teilsequenzen homologen DNA-Sequenzen, einem Verfahren zur Herstellung von Pflanzen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen, sowie die derart hergestellte Pflanze selbst.

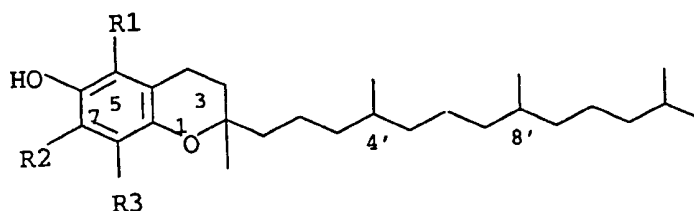
20 Ein wichtiges Ziel pflanzenmolekulargenetischer Arbeiten ist bisher die Erzeugung von Pflanzen mit erhöhtem Gehalt an Zuckern, Enzymen und Aminosäuren. Wirtschaftlich interessant ist jedoch auch die Entwicklung von Pflanzen mit erhöhtem Gehalt an Vitaminen, wie z.B. der Erhöhung des Tocopherol- und Tocotrienolgehal-

25 tes.

Die in der Natur vorkommenden acht Verbindungen mit Vitamin E-Aktivität sind Derivate des 6-Chromanols (Ullmann's Encyclopedia of Industrial Chemistry, Vol. A 27 (1996), VCH Verlagsgesellschaft, Chapter 4., 478-488, Vitamin E). Die erste Gruppe (1a-d)

30 stammt von Tocopherol ab, die zweite Gruppe besteht aus Derivaten des Tocotrienols (2a- d):

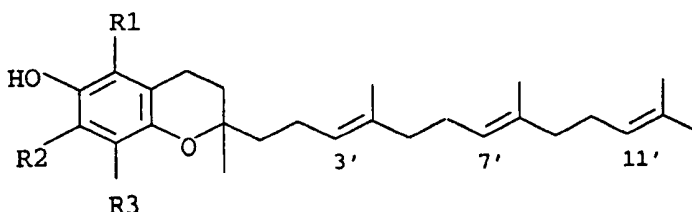
35



- 40 1a, α -Tocopherol: $R^1 = R^2 = R^3 = \text{CH}_3$
 1b, β -Tocopherol [148-03-8]: $R^1 = R^3 = \text{CH}_3$, $R^2 = \text{H}$
 1c, γ -Tocopherol [54-28-4]: $R^1 = \text{H}$, $R^2 = R^3 = \text{CH}_3$
 1d, δ -Tocopherol [119-13-1]: $R^1 = R^2 = \text{H}$, $R^3 = \text{CH}_3$

45

5



- 2a, α -Tocotrienol [1721-51-3]: $R^1 = R^2 = R^3 = \text{CH}_3$
 2b, β -Tocotrienol [490-23-3]: $R^1 = R^3 = \text{CH}_3$, $R^2 = \text{H}$
 10 2c, γ -Tocotrienol [14101-61-2]: $R^1 = \text{H}$, $R^2 = R^3 = \text{CH}_3$
 2d, δ -Tocotrienol [25612-59-3]: $R^1 = R^2 = \text{H}$, $R^3 = \text{CH}_3$

Wirtschaftlich große Bedeutung besitzt α -Tocopherol.

- 15 Der Entwicklung von Kulturpflanzen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen durch klassische Züchtungsmethoden sind Grenzen gesetzt.

- Eine sinnvolle Alternative ist das gentechnische Vorgehen,
 20 beispielsweise die für die Tocopherol Syntheseleistung kodierenden, essentiellen Biosynthesegene zu isolieren und in Kulturpflanzen gezielt zu übertragen. Dieses Verfahren setzt voraus, daß die Biosynthese und deren Regulation bekannt ist und daß Gene, die die Biosyntheseleistung beeinflussen, identifiziert
 25 werden.

Isoprenoide oder Terpenoide bestehen aus verschiedenen Klassen lipidlöslicher Moleküle und werden teilweise oder vollständig aus C_5 -Isopren-Einheiten gebildet. Reine Prenyllipide (z.B.

- 30 Carotinoide) bestehen aus C-Gerüsten, die ausschließlich auf Isopren-Einheiten zurückgehen, während gemischte Prenyllipide (z.B. Chlorophylle, Tocopherole und Vitamin K) eine Isoprenoid-Seitenkette besitzen, die mit einem aromatischen Kern verbunden ist.

- 35 Ausgangspunkt der Biosynthese von Prenyllipiden sind 3 x Acetyl-CoA Einheiten, die über β -Hydroxymethylglutaryl-CoA (HMG-CoA) und Mevalonat in die Ausgangs-Isopren-Einheit (C_5), dem Isopentenylpyrophosphat (IPP), umgewandelt werden. Kürzlich wurde durch *in vivo* Fütterungsexperimente mit C^{13} gezeigt, daß in verschiedenen
 40 Eubakterien, Grünalgen und pflanzlichen Chloroplasten ein Mevalonat-unabhängiger Weg zur Bildung von IPP beschritten wird. Dabei werden Hydroxyethylthiamin, das durch Decarboxylierung von Pyruvat entsteht, und Glycerinaldehyd-3-Phosphat (3-GAP) in einer durch die 1-Deoxy-D-Xylulose-5-Phosphat Synthase vermittelten
 45 "Transketolase"-Reaktion zunächst in 1-Deoxy-D-Xylulose-5-phosphat umgewandelt (Lange et al, 1998; Schwender et al, 1997; Arigoni et al, 1997; Lichtenthaler et al, 1997; Sprenger et

al, 1997). Dieses wird dann durch eine intramolekulare Umordnung in 2-C-Methyl-D-Erythritol-4-Phosphat und im weiteren zu IPP umgesetzt (Arigoni et al, 1997; Zeidler et al, 1998). Biochemische Daten deuten darauf hin, daß der Mevalonat-Weg im Zytosol operiert und zur Bildung von Phytosterolen führt. Das Antibiotikum Mevinolin, ein spezifischer Inhibitor der Mevalonat-Bildung, führt lediglich zur Inhibition der Sterol-Biosynthese im Zytoplasma, während die Prenyllipid-Bildung in den Plastiden unbeeinflusst ist (Bach & Lichtenthaler, 1993). Der Mevalonat-unabhängige Weg ist dagegen plastidär lokalisiert und führt vornehmlich zur Bildung von Carotinoiden und plastidären Prenyllipiden (Schwender et al, 1997; Arigoni et al, 1997).

IPP steht im Gleichgewicht mit seinem Isomer, dem Dimethylallyl Pyrophosphat (DMAPP). Eine Kondensation von IPP mit DMAPP in Kopf-Schwanz Anlagerung ergibt das Monoterpen (C₁₀) Geranyl-Pyrophosphat (GPP). Die Addition von weiteren IPP Einheiten führt zum Sesquiterpen (C₁₅) Farnesyl-Pyrophosphat (FPP) und zum Diterpen (C₂₀) Geranyl-Geranyl-Pyrophosphat (GGPP). Die Verknüpfung zweier GGPP Moleküle führt zur Bildung der C₄₀-Vorläufer für Carotinoide.

Bei gemischten Prenyllipiden ist die Isopren-Seitenkette verschiedener Länge mit Nicht-Isopren Ringen verbunden wie beispielsweise ein Porphyrin-Ring bei Chlorophyll a und b. Die Chlorophylle und Phylloquinone enthalten eine C₂₀ Phytyl-Kette, in der nur die erste Isopren-Einheit eine Doppelbindung enthält. GGPP wird durch die Geranylgeranyl-Pyrophosphat-Oxidoreduktase (GGPPOR) zum Phytyl-Pyrophosphat (PPP) umgeformt, dem Ausgangsstoff für die weitere Bildung von Tocopherolen.

Bei den Ringstrukturen der gemischten Prenyllipide, die zur Bildung der Vitamine E und K führen, handelt es sich um Quinone, deren Ausgangsmetabolite aus dem Shikimat-Weg stammen. Die aromatischen Aminosäuren Phenylalanin bzw. Tyrosin werden in Hydroxyphenyl-Pyruvat umgewandelt, welches durch Dioxygenierung in Homogentisinsäure überführt wird. Das Chorismat wird ausgehend von Erythrose-4-Phosphat und Phosphoenolpyruvat (PEP) durch deren Kondensation zu 3-deoxy-D-Arabino-heptulosonat-7-Phosphat (DAHP) über die Zwischenstufen des Shikimatweges 3'-Dehydroquinat, 3'-Dehydroshikimat, Shikimat, Shikimat-3-Phosphat und 5'-Enolpyruvylshikimat-3-Phosphat gebildet. Dabei wird das Erythrose-4-Phosphat vom Calvinzyklus gebildet und das PEP von der Glykolyse bereitgestellt. Die oben beschriebene Homogentisinsäure wird anschließend an Phytylpyrophosphat (PPP) bzw. Geranylgeranylpyrophosphat gebunden, um die Vorläufer von α -Tocopherol und α -Tocotrienol, das 2-Methyl-6-phytylhydrochinon bzw. das

2-Methyl-6-geranylgeranylhydrochinon zu bilden. Durch Methylierungsschritte mit S-Adenosylmethionin als Methyl-Gruppen-Donor entsteht zunächst 2,3-Dimethyl-6-phytylquinol, dann durch Zyklisierung α -Tocopherol und durch nochmalige Methylierung α -Tocopherol (Richter, Biochemie der Pflanzen, Georg Thieme Verlag Stuttgart, 1996).

In der Literatur finden sich Beispiele die zeigen, daß die Manipulation eines Enzyms den Metabolit-Fluß direktional beeinflussen kann. In Experimenten mit einer veränderten Expression der Phytoen Synthase, welche zwei GGPP-Moleküle zu 15-cis-Phytoen miteinander verknüpft, konnte ein direkter Einfluß auf die Carotinoid-Mengen dieser transgenen Tomatenpflanzen gemessen werden (Fray und Grierson, Plant Mol.Biol.22(4), 589-602(1993); Fray et al., Plant J., 8, 693-701(1995)). Wie zu erwarten, zeigen transgene Tabakpflanzen mit verringerten Mengen an Phenylalanin-Ammonium Lyase reduzierte Phenylpropanoid-Mengen. Das Enzym Phenylalanin-Ammonium Lyase katalysiert den Abbau von Phenylalanin, entzieht es also der Phenylpropanoid-Biosynthese (Bate et al., Proc. Natl. Acad. Sci USA 91 (16): 7608-7612 (1994); Howles et al., Plant Physiol. 112. 1617-1624(1996)).

Über die Erhöhung des Metabolitflusses zur Steigerung des Tocopherol- bzw. Tocotrienolgehaltes in Pflanzen durch Überexpression einzelner Biosynthesegene ist bisher wenig bekannt. Lediglich WO 97/27285 beschreibt eine Modifikation des Tocopherol-Gehaltes durch verstärkte Expression bzw. durch Herunterregulation des Enzyms p-Hydroxyphenylpyruvatdioxygenase (HPPD). WO 99/04622 beschreibt eine Gensequenz codierend für eine γ -Tocopherolmethyltransferase aus einem photosynthetisch aktiven Organismus. WO 99/23231 zeigt, daß die Expression einer Geranylgeranyl-Reductase in transgenen Pflanzen eine gesteigerte Tocopherolbiosynthese zur Folge hat.

Aufgabe der vorliegenden Erfindung war die Entwicklung einer transgenen Pflanze mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen.

Die Aufgabe wurden überraschenderweise gelöst durch die Überexpression eines 2-Methyl-6-phytylhydrochinon-methyltransferase Gens in Pflanzen.

Zu diesem Zweck wurde in transgenen Pflanzen die Aktivität der 2-Methyl-6-phytylhydrochinon-methyltransferase (MPMT) durch Überexpression des MPMT-Gens aus Synechocystis spec. PCC 6803 er-

hört. Dies kann prinzipiell durch Expression homologer oder heterologer MPMT-Gene erreicht werden.

In Beispiel 2 wird erstmals die Klonierung einer MPMT-DNA-Sequenz
5 (SEQ-ID Nr. 1) aus *Synechocystis spec.* PCC 6803 beschrieben. Um eine Plastidenlokalisation zu gewährleisten wird der MPMT-Nukleotidsequenz aus *Synechocystis* eine Transitsignalsequenz (Abb. 3, Abb. 4) vorangestellt. Auch geeignet als Expressionskassette ist eine DNA-Sequenz, die für ein MPMT-Gen codiert, das mit
10 SEQ-ID Nr. 1 hybridisiert, bzw. zur Gesamtsequenz oder zu Teilsequenzen homolog ist und das aus anderen Organismen bzw. aus Pflanzen stammt.

Das durch die zusätzliche Expression des MPMT-Gens nun vermehrt
15 zur Verfügung stehende 2,3-Dimethyl-6-phytylhydrochinon wird weiter in Richtung Tocopherole und Tocotrienol umgesetzt (Abbildung 1).

Die Herstellung der transgenen Pflanzen erfolgt durch Transforma-
20 tion der Pflanzen mit einem das MPMT-Gen enthaltenden Konstrukt. Als Modellpflanzen für die Produktion von Tocopherolen und Tocotrienolen wurden *Arabidopsis thaliana*, *Brassica napus* und *Nicotiana tabacum* eingesetzt.

25 Messungen an MPMT-*Synechocystis* knock out Mutanten ergaben bezüglich des Gehaltes an Tocopherolen und Tocotrienolen eine drastische Abnahme. Dies belegt den direkten Einfluß der plastidären pflanzlichen MPMT auf die Synthese von Tocopherolen und Tocotrienolen.

30 Gegenstand der Erfindung ist die Verwendung einer DNA-Sequenz SEQ-ID Nr. 1 aus *Synechocystis spec.* PCC 6803, die für eine MPMT oder deren funktionelle Äquivalente kodiert, zur Herstellung einer Pflanze mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen. Die Nukleinsäuresequenz kann dabei z.B. eine DNA- oder cDNA-
35 Sequenz sein. Zur Insertion in eine Expressionskassette geeignete kodierende Sequenzen sind beispielsweise solche, die für eine MPMT kodieren und die dem Wirt die Fähigkeit zur Überproduktion von Tocopherolen und Tocotrienolen verleihen.

40 Die Expressionskassetten beinhalten außerdem regulative Nukleinsäuresequenzen, welche die Expression der kodierenden Sequenz in der Wirtszelle steuern. Gemäß einer bevorzugten Ausführungsform umfaßt eine Expressionskassette stromaufwärts, d.h. am 5'-Ende
45 der kodierenden Sequenz, einen Promotor und stromabwärts, d.h. am 3'-Ende, ein Polyadenylierungssignal und gegebenenfalls weitere regulatorische Elemente, welche mit der dazwischenliegenden ko-

dierenden Sequenz für das MPMT-Gen operativ verknüpft sind. Unter einer operativen Verknüpfung versteht man die sequenzielle Anordnung von Promotor, kodierender Sequenz, Terminator und ggf. weiterer regulativer Elemente derart, daß jedes der regulativen Elemente seine Funktion bei der Expression der kodierenden Sequenz bestimmungsgemäß erfüllen kann. Die zur operativen Verknüpfung bevorzugten aber nicht darauf beschränkten Sequenzen sind Targeting-Sequenzen zur Gewährleistung der subzellulären Lokalisation im Apoplasten, in der Vakuole, in Plastiden, im Mitochondrium, im Endoplasmatischen Retikulum (ER), im Zellkern, in Ölkörperchen oder anderen Kompartimenten und Translationsverstärker wie die 5'-Führungssequenz aus dem Tabak-Mosaik-Virus (Gallie et al., Nucl. Acids Res. 15 (1987), 8693 -8711).

15 Beispielhaft kann die pflanzliche Expressionskassette in ein Derivat des Transformationsvektors pBin-19 mit 35s Promotor (Bevan, M., Nucleic Acids Research 12: 8711-8721 (1984)) eingebaut werden. Abbildung 4 zeigt ein Derivat des Transformationsvektors pBin -19 mit samenspezifischem Legumin B4-Promotor.

20 Als Promotoren der Expressionskassette ist grundsätzlich jeder Promotor geeignet, der die Expression von Fremdgenen in Pflanzen steuern kann. Vorzugsweise verwendet man insbesondere einen pflanzlichen Promotor oder einen Promotor, der einem Pflanzenvirus entstammt. Insbesondere bevorzugt ist der CaMV 35S-Promotor aus dem Blumenkohl-Mosaik-Virus (Franck et al., Cell 21 (1980), 285 - 294). Dieser Promotor enthält bekanntlich unterschiedliche Erkennungssequenzen für transkriptionale Effektoren, die in ihrer Gesamtheit zu einer permanenten und konstitutiven Expression des
25 eingeführten Gens führen (Benfey et al., EMBO J. 8 (1989), 2195-2202).

Die Expressionskassette kann auch einen chemisch induzierbaren Promotor enthalten, durch den die Expression des exogenen MPMT-Gens in der Pflanze zu einem bestimmten Zeitpunkt gesteuert werden kann. Derartige Promotoren wie z.B. der PRP1-Promotor (Ward et al., Plant. Mol. Biol. 22 (1993), 361-366), ein durch Salizylsäure induzierbarer Promotor (WO 95/19443), ein durch Benzenesulfonamid-induzierbarer (EP-A 388186), ein durch Tetrazyklin-induzierbarer (Gatz et al., (1992) Plant J. 2, 397-404), ein durch Abscisinsäure-induzierbarer (EP-A 335528) bzw. ein durch Ethanol- oder Cyclohexanon-induzierbarer (WO 93/21334) Promotor können u.a. verwendet werden.

45 Weiterhin sind insbesondere solche Promotoren bevorzugt, die die Expression in Geweben oder Pflanzenteilen sicherstellen, in denen beispielsweise die Biosynthese von Tocopherol bzw. dessen Vorstu-

fen stattfindet. Insbesondere zu nennen sind Promotoren, die eine blattspezifische Expression gewährleisten. Zu nennen sind der Promotor der cytosolischen FBPase aus Kartoffel oder der ST-LSI Promotor aus Kartoffel (Stockhaus et al., EMBO J. 8 (1989), 5 2445 - 245).

Mit Hilfe eines samenspezifischen Promotors konnte ein Fremdprotein stabil bis zu einem Anteil von 0,67 % des gesamten löslichen Samenproteins in den Samen transgener Tabakpflanzen exprimiert werden (Fiedler und Conrad, Bio/Technology 10 (1995), 1090-1094). Die Expressionskassette kann daher beispielsweise einen samenspezifischen Promotor (bevorzugt den Phaseolin-Promotor (US 5504200), den USP- (Baumlein, H. et al., Mol. Gen. Genet. (1991) 225 (3), 459 - 467) oder LEB4-Promotor (Fiedler und Conrad, 1995)), das LEB4-Signalpeptid, das zu exprimierende Gen und ein ER-Retentionssignal enthalten.

Die Herstellung einer Expressionskassette erfolgt durch Fusion eines geeigneten Promotors mit einer geeigneten MPMT-DNA Sequenz und vorzugsweise einer zwischen Promotor und MPMT-DNA-Sequenz inserierten DNA, die für ein chloroplastenspezifisches Transitpeptid kodiert, sowie einem Polyadenylierungssignal nach gängigen Rekombinations- und Klonierungstechniken, wie sie beispielsweise in T. Maniatis, E.F. Fritsch und J. Sambrook, Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1989) sowie in T.J. Silhavy, M.L. Berman und L.W. Enquist, Experiments with Gene Fusions, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY (1984) und in Ausubel, F.M. et al., Current Protocols in Molecular Biology, Greene Publishing Assoc. and Wiley-Interscience (1987) beschrieben sind.

Insbesondere bevorzugt sind Sequenzen, die ein Targeting in den Plastiden gewährleisten.

Es können auch Expressionskassetten verwendet werden, deren DNA-Sequenz für ein MPMT-Fusionsprotein kodiert, wobei ein Teil des Fusionsproteins ein Transitpeptid ist, das die Translokation des Polypeptides steuert. Bevorzugt sind für die Chloroplasten spezifische Transitpeptide, welche nach Translokation des MPMT-Gens in die Chloroplasten vom MPMT-Teil enzymatisch abgespalten werden. Insbesondere bevorzugt ist das Transitpeptid, das von der plastidären *Nicotiana tabacum* Transketolase oder einem anderen Transitpeptid (z.B. dem Transitpeptid der kleinen Untereinheit der Rubisco oder der Ferredoxin NADP Oxidoreduktase) oder dessen funktionellem Äquivalent abgeleitet ist.

Besonders bevorzugt sind DNA-Sequenzen von drei Kassetten des Plastiden-Transitpeptids der plastidären Transketolase aus Tabak in drei Leserastern als KpnI/BamHI Fragmente mit einem ATG-Codon in der NcoI Schnittstelle:

5

pTP09

```
KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTGTC
CCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCTTCTTCTCTCACTTTTTCCGGCCTTAA
10 ATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCCGCCGCCGTCG
TAAGGTCACCGGCGATTTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAAAGTGAAGTGCAGGGA
TCC_BamHI
```

pTP10

15

```
KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTGTC
CCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCTTCTTCTCTCACTTTTTCCGGCCTTAA
ATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCCGCCGCCGTCG
TAAGGTCACCGGCGATTTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAAAGTGAAGTGCAGTGC
20 GATCC_BamHI
```

pTP11

```
KpnI_GGTACCATGGCGTCTTCTTCTTCTCTCACTCTCTCTCAAGCTATCCTCTCTCGTTCTGTC
25 CCTCGCCATGGCTCTGCCTCTTCTTCTCAACTTTCCCTTCTTCTCTCACTTTTTCCGGCCTTAA
ATCCAATCCCAATATCACCACCTCCCGCCGCCGTACTCCTTCCTCCGCCGCCGCCGCCGCCGTCG
TAAGGTCACCGGCGATTTCGTGCCTCAGCTGCAACCGAAACCATAGAGAAAAGTGAAGTGCAGGGG
ATCC_BamHI
```

30 Die inserierte Nukleotid-Sequenz kodierend für eine MPMT kann synthetisch hergestellt oder natürlich gewonnen sein oder eine Mischung aus synthetischen und natürlichen DNA-Bestandteilen enthalten, sowie aus verschiedenen heterologen MPMT-Genabschnitten verschiedener Organismen bestehen. Im allgemeinen werden synthetische Nukleotid-Sequenzen mit Kodons erzeugt, die von Pflanzen bevorzugt werden. Diese von Pflanzen bevorzugten Kodons können aus Kodons mit der höchsten Proteinhäufigkeit bestimmt werden, die in den meisten interessanten Pflanzenspezies exprimiert werden. Bei der Präparation einer Expressionskassette können verschiedene DNA-Fragmente manipuliert werden, um eine Nukleotid-Sequenz zu erhalten, die zweckmäßigerweise in der korrekten Richtung liest und die mit einem korrekten Leseraster ausgestattet ist. Für die Verbindung der DNA-Fragmente miteinander können an die Fragmente Adaptoren oder Linker angesetzt werden.

45

Zweckmäßigerweise können die Promotor- und die Terminator-Regionen in Transkriptionsrichtung mit einem Linker oder Polylinker, der eine oder mehrere Restriktionsstellen für die Insertion dieser Sequenz enthält, versehen werden. In der Regel hat der Linker 5 1 bis 10, meistens 1 bis 8, vorzugsweise 2 bis 6 Restriktionsstellen. Im allgemeinen hat der Linker innerhalb der regulatorischen Bereiche eine Größe von weniger als 100 bp, häufig weniger als 60 bp, mindestens jedoch 5 bp. Der Promotor kann sowohl nativ bzw. homolog als auch fremdartig bzw. heterolog zur Wirtspflanze 10 sein. Die Expressionskassette beinhaltet in der 5'-3'-Transkriptionsrichtung den Promotor, eine DNA-Sequenz die für ein MPMT-Gen codiert und eine Region für die transkriptionale Termination. Verschiedene Terminationsbereiche sind gegeneinander beliebig austauschbar.

15 Ferner können Manipulationen, die passende Restriktionsschnittstellen bereitstellen oder die überflüssige DNA oder Restriktionsschnittstellen entfernen, eingesetzt werden. Wo Insertionen, Deletionen oder Substitutionen wie z.B. Transitionen und Transversionen in Frage kommen, können *in vitro*-Mutagenese, "primerre- 20 pair", Restriktion oder Ligation verwendet werden. Bei geeigneten Manipulationen, wie z.B. Restriktion, "chewing-back" oder Auffüllen von Überhängen für "bluntends", können komplementäre Enden der Fragmente für die Ligation zur Verfügung gestellt werden.

25 Bevorzugte Polyadenylierungssignale sind pflanzliche Polyadenylierungssignale, vorzugsweise solche, die im wesentlichen T-DNA-Polyadenylierungssignale aus *Agrobacterium tumefaciens*, insbesondere des Gens 3 der T-DNA (Octopin Synthase) des Ti-Plasmids 30 pTiACH5 entsprechen (Gielen et al., EMBO J. 3 (1984), 835 ff) oder funktionelle Äquivalente.

Vorzugsweise wird die fusionierte Expressionskassette, die für ein MPMT-Gen kodiert, in einen Vektor, beispielsweise pBin19, 35 kloniert, der geeignet ist, *Agrobacterium tumefaciens* zu transformieren. Mit einem solchen Vektor transformierte Agrobakterien können dann in bekannter Weise zur Transformation von Pflanzen, insbesondere von Kulturpflanzen, wie z.B. von Tabakpflanzen, verwendet werden, indem beispielsweise verwundete Blätter oder 40 Blattstücke in einer Agrobakterienlösung gebadet und anschließend in geeigneten Medien kultiviert werden. Die Transformation von Pflanzen durch Agrobakterien ist unter anderem bekannt aus F.F. White, Vectors for Gene Transfer in Higher Plants; in Transgenic Plants, Vol. 1, Engineering and Utilization, herausgegeben von 45 S.D. Kung und R. Wu, Academic Press, 1993, S. 15 - 38. Aus den transformierten Zellen der verwundeten Blätter bzw. Blattstücke können in bekannter Weise transgene Pflanzen regeneriert werden,

die ein in die Expressionskassette integriertes Gen für die Expression eines MPMT-Gens enthalten.

Zur Transformation einer Wirtspflanze mit einer für eine MPMT kodierenden DNA wird eine Expressionskassette als Insertion in einen rekombinanten Vektor eingebaut, dessen Vektor-DNA zusätzliche funktionelle Regulationssignale, beispielsweise Sequenzen für Replikation oder Integration enthält. Geeignete Vektoren sind unter anderem in "Methods in Plant Molecular Biology and Biotechnology" (CRC Press), Kap. 6/7, S. 71 - 119 (1993) beschrieben.

Unter Verwendung der oben zitierten Rekombinations- und Klonierungstechniken können die Expressionskassetten in geeignete Vektoren kloniert werden, die ihre Vermehrung, beispielsweise in *E. coli*, ermöglichen. Geeignete Klonierungsvektoren sind u.a. pBR332, pUC-Serien, M13mp-Serien und pACYC184. Besonders geeignet sind binäre Vektoren, die sowohl in *E. coli* als auch in Agrobakterien replizieren können.

Ein weiterer Gegenstand der Erfindung betrifft die Verwendung einer Expressionskassette enthaltend eine DNA-Sequenz SEQ-ID Nr. 1 oder eine mit dieser hybridisierende DNA-Sequenz zur Transformation von Pflanzen, -zellen, -geweben oder Pflanzenteilen. Vorzugsweise ist Ziel der Verwendung die Erhöhung des Gehaltes an Tocopherolen und Tocotrienolen der Pflanze.

Dabei kann je nach Wahl des Promotors die Expression spezifisch in den Blättern, in den Samen, Blütenblättern oder anderen Teilen der Pflanze erfolgen. Solche transgenen Pflanzen, deren Vermehrungsgut, sowie deren Pflanzenzellen, -gewebe oder -teile sind ein weiterer Gegenstand der vorliegenden Erfindung.

Die Expressionskassette kann darüberhinaus auch zur Transformation von Bakterien, Cyanobakterien, Hefen, filamentösen Pilzen und Algen mit dem Ziel einer Erhöhung des Gehaltes an Tocopherolen und Tocotrienolen eingesetzt werden.

Die Übertragung von Fremdgenen in das Genom einer Pflanze wird als Transformation bezeichnet. Es werden dabei die beschriebenen Methoden zur Transformation und Regeneration von Pflanzen aus Pflanzengeweben oder Pflanzenzellen zur transienten oder stabilen Transformation genutzt. Geeignete Methoden sind die Protoplastentransformation durch Polyethylenglykol-induzierte DNA-Aufnahme, das biolistische Verfahren mit der Genkanone - die sogenannte particle bombardment Methode, die Elektroporation, die Inkubation trockener Embryonen in DNA-haltiger Lösung, die Mikroinjektion



(84) **Bestimmungsstaaten** (*regional*): ARIPO-Patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI-Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Veröffentlicht:

— Mit internationalem Recherchenbericht.

— Vor Ablauf der für Änderungen der Ansprüche geltenden Frist; Veröffentlichung wird wiederholt, falls Änderungen eintreffen.

Zur Erklärung der Zweibuchstaben-Codes, und der anderen Abkürzungen wird auf die Erklärungen ("Guidance Notes on Codes and Abbreviations") am Anfang jeder regulären Ausgabe der PCT-Gazette verwiesen.

und der durch Agrobacterium vermittelte Gentransfer. Die genannten Verfahren sind beispielsweise in B. Jenes et al., *Techniques for Gene Transfer*, in: *Transgenic Plants*, Vol. 1, Engineering and Utilization, herausgegeben von S.D. Kung und R. Wu, Academic Press (1993), 128 - 143 sowie in Potrykus, *Annu. Rev. Plant Physiol. Plant Molec. Biol.* 42 (1991), 205 - 225) beschrieben. Vorzugsweise wird das zu exprimierende Konstrukt in einen Vektor kloniert, der geeignet ist, Agrobacterium tumefaciens zu transformieren, beispielsweise pBin19 (Bevan et al., *Nucl. Acids Res.* 12 (1984), 8711).

Mit einer Expressionskassette transformierte Agrobakterien können ebenfalls in bekannter Weise zur Transformation von Pflanzen, insbesondere von Kulturpflanzen, wie Getreide, Mais, Hafer, Soja, Reis, Baumwolle, Zuckerrübe, Canola, Sonnenblume, Flachs, Hanf, Kartoffel, Tabak, Tomate, Raps, Alfalfa, Salat und den verschiedenen Baum-, Nuß- und Weinspezies, verwendet werden, z.B. indem verwundete Blätter oder Blattstücke in einer Agrobakterienlösung gebadet und anschließend in geeigneten Medien kultiviert werden.

Funktionell äquivalente Sequenzen, die für ein MPMT-Gen kodieren, sind solche Sequenzen, welche trotz abweichender Nukleotidsequenz noch die gewünschten Funktionen besitzen. Funktionelle Äquivalente umfassen somit natürlich vorkommende Varianten der hierin beschriebenen Sequenzen sowie künstliche, z.B. durch chemische Synthese erhaltene, an den Kodon-Gebrauch einer Pflanze angepasste, künstliche Nukleotid-Sequenzen.

Unter einem funktionellen Äquivalent versteht man insbesondere auch natürliche oder künstliche Mutationen einer ursprünglich isolierten für eine MPMT kodierende Sequenz, welche weiterhin die gewünschte Funktion zeigen. Mutationen umfassen Substitutionen, Additionen, Deletionen, Vertauschungen oder Insertionen eines oder mehrerer Nukleotidreste. Somit werden beispielsweise auch solche Nukleotidsequenzen durch die vorliegende Erfindung mit umfaßt, welche man durch Modifikation der MPMT-Nukleotidsequenz erhält. Ziel einer solchen Modifikation kann z.B. die weitere Eingrenzung der darin enthaltenen kodierenden Sequenz oder z.B. auch die Einfügung weiterer Restriktionsenzym-Schnittstellen sein.

Beispiel 8 beschreibt einen Deletionsklon des MPMT-Gens, siehe SEQ-ID Nr. 7)

Funktionelle Äquivalente sind auch solche Varianten, deren Funktion, verglichen mit dem Ausgangsgen bzw. Genfragment, abgeschwächt oder verstärkt ist.

- Außerdem sind artifizielle DNA-Sequenzen geeignet, solange sie, wie oben beschrieben, die gewünschte Eigenschaft beispielsweise der Erhöhung des Tocopherol-Gehaltes in der Pflanze durch Überexpression eines MPMT-Gens in Kulturpflanzen vermitteln. Solche
- 5 artifiziellen DNA-Sequenzen können beispielsweise durch Rückübersetzung mittels Molecular Modelling konstruierter Proteine, die MPMT-Aktivität aufweisen oder durch *in vitro*-Selektion ermittelt werden. Besonders geeignet sind kodierende DNA-Sequenzen, die durch Rückübersetzung einer Polypeptidsequenz gemäß der für die
- 10 Wirtspflanze spezifischen Kodon-Nutzung erhalten wurden. Die spezifische Kodon-Nutzung kann ein mit pflanzen genetischen Methoden vertrauter Fachmann durch Computerauswertungen anderer, bekannter Gene der zu transformierenden Pflanze leicht ermitteln.
- 15 Als weitere geeignete äquivalente Nukleinsäure-Sequenzen sind zu nennen Sequenzen, welche für Fusionsproteine kodieren, wobei Bestandteil des Fusionsproteins ein MPMT-Polypeptid oder ein funktionell äquivalenter Teil davon ist. Der zweite Teil des Fusionsproteins kann z.B. ein weiteres Polypeptid mit enzymatischer
- 20 Aktivität sein oder eine antigene Polypeptidsequenz mit deren Hilfe ein Nachweis auf MPMT-Expression möglich ist (z.B. myc-tag oder his-tag). Bevorzugt handelt es sich dabei jedoch um eine regulative Proteinsequenz, wie z.B. ein Transitpeptid, das das MPMT-Protein in die Plastiden leitet.
- 25 Erhöhung des Gehaltes an Tocopherolen und Tocotrienolen bedeutet im Rahmen der vorliegenden Erfindung die künstlich erworbene Fähigkeit einer erhöhten Biosyntheseleistung dieser Verbindungen durch funktionelle Überexpression eines MPMT-Gens SEQ-ID Nr. 1
- 30 oder SEQ-ID Nr. 7 in der Pflanze gegenüber der nicht gentechnisch modifizierten Pflanze für die Dauer mindestens einer Pflanzengeneration.

Dabei kann sowohl der Gehalt an Tocopherolen und Tocotrienolen

35 gesteigert werden. Vorzugsweise wird der Gehalt an Tocopherolen gesteigert. Aber es ist auch möglich unter bestimmten Bedingungen vorzugsweise den Gehalt an Tocotrienolen zu steigern.

Der Biosyntheseort von Tocopherolen beispielsweise ist unter anderem das Blattgewebe, so daß eine blattspezifische Expression

40 des MPMT-Gens sinnvoll ist. Es ist jedoch naheliegend, daß die Tocopherol-Biosynthese nicht auf das Blattgewebe beschränkt sein muß, sondern auch in allen übrigen Teilen der Pflanze - besonders in fetthaltigen Samen - gewebespezifisch erfolgen kann.

Darüberhinaus ist eine konstitutive Expression des exogenen MPMT-Gens von Vorteil. Andererseits kann aber auch eine induzierbare Expression wünschenswert erscheinen.

- 5 Die Wirksamkeit der Expression des transgen exprimierten MPMT-Gens kann beispielsweise *in vitro* durch Sproßmeristemvermehrung ermittelt werden. Zudem kann eine in Art und Höhe veränderte Expression des MPMT-Gens und deren Auswirkung auf die Tocopherol-Biosyntheseleistung an Testpflanzen in Gewächshausversuchen getestet werden.

- Gegenstand der Erfindung sind außerdem transgene Pflanzen, transformiert mit einer Expressionskassette enthaltend die Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder eine mit dieser hybridisierende bzw. zur Gesamtsequenz oder zu Teilsequenzen homologen DNA-Sequenz, sowie transgene Zellen, Gewebe, Teile und Vermehrungsgut solcher Pflanzen. Besonders bevorzugt sind dabei transgene Kulturpflanzen, wie z.B. Gerste, Weizen, Roggen, Mais, Hafer, Soja, Reis, Baumwolle, Zuckerrübe, Canola, Sonnenblume, Flachs, Hanf, Kartoffel, Tabak, Tomate, Raps, Alfalfa, Tagetes, Salat und die verschiedenen Baum-, Nuß- und Weinspezies.

Pflanzen im Sinne der Erfindung sind mono- und dikotyle Pflanzen.

- 25 Gegenstand der Erfindung sind weiterhin photosynthetisch aktive Organismen transformiert mit einer Expressionskassette enthaltend die Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder eine mit dieser hybridisierende bzw. zur Gesamtsequenz oder zu Teilsequenzen homologen DNA-Sequenz. Photosynthetisch aktive Organismen sind neben Pflanzen, beispielsweise Cyanobakterien, Moose und Algen.

- Da es sich bei diesem Biosyntheseweg um einen ausschließlich plastidär-lokalisierten Stoffwechselweg handelt, bietet er optimale Targetenzyme für die Entwicklung von Inhibitoren. Da sich nach heutigem Stand der Technik kein mit der *Synechocystis* MPMT identisches oder ähnliches Enzym in humanen und tierischen Organismen befindet, ist davon auszugehen, daß Inhibitoren sehr spezifisch auf Pflanzen wirken sollten.

- 40 Wie bereits erwähnt ist die MPMT ein potentiell Target für Herbizide. Um effiziente Hemmstoffe der MPMT finden zu können, ist es notwendig, geeignete Testsysteme, mit denen Inhibitor-Enzym-Bindungsstudien durchgeführt werden können, zur Verfügung zu stellen. Hierzu wird beispielsweise die komplette cDNA-Sequenz der MPMT aus *Synechocystis* in einen Expressionsvektor (pQE, Qia-gen) kloniert und in *E. coli* überexprimiert.

Das mit Hilfe der erfindungsgemäßen Expressionskassette expri-
mierte MPMT-Protein eignet sich besonders zur Auffindung von für
die MPMT spezifischen Hemmstoffen.

- 5 Dazu kann die MPMT beispielsweise in einem Enzymtest eingesetzt
werden, bei dem die Aktivität der MPMT in An- und Abwesenheit
des zu testenden Wirkstoffs ermittelt wird. Aus dem Vergleich
der beiden Aktivitätsbestimmungen läßt sich eine qualitative und
quantitative Aussage über das Hemmverhalten des zu testenden
10 Wirkstoffes machen.

- Mit Hilfe des erfindungsgemäßen Testsystems kann eine Vielzahl
von chemischen Verbindungen schnell und einfach auf herbizide
Eigenschaften überprüft werden. Das Verfahren gestattet es,
15 reproduzierbar aus einer großen Anzahl von Substanzen gezielt
solche mit großer Wirkstärke auszuwählen, um mit diesen Sub-
stanzen anschließend weitere, dem Fachmann geläufige vertiefte
Prüfungen durchzuführen.
- 20 Ein weiterer Gegenstand der Erfindung sind Herbizide, die mit dem
oben beschriebenen Testsystem identifizierbar sind.

- Durch Überexpression der für eine MPMT kodierenden Gensequenz
SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 in einer Pflanze wird eine erhöhte
25 Resistenz gegenüber Inhibitoren der MPMT erreicht. Die derart
hergestellten transgenen Pflanzen sind ebenfalls Gegenstand der
Erfindung.

- Das unter Verwendung der DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID
30 Nr. 7 hergestellte MPMT-Protein eignet sich auch zur Durchführung
von Biotransformationen zur Bereitstellung größerer Mengen
2,3-Dimethyl-6-phytylhydrochinon. Dabei wird 2-Methyl-6-phytylhy-
drochinon in Gegenwart des Enzyms MPMT und des Cosubstrats S-Ade-
nosyl-L-Methionin zu 2,3-Dimethyl-6-phytylhydrochinon umgesetzt.
- 35 Die Biotransformation läßt sich prinzipiell mit ganzen Zellen,
die das Enzym MPMT exprimieren oder Zellextrakten aus diesen Zel-
len oder aber mit aufgereinigter oder hochreiner MPMT in Gegen-
wart von S-Adenosyl-L-Methionin durchführen.

- 40 Weitere Gegenstände der Erfindung sind:

- Verfahren zur Transformation einer Pflanze dadurch gekenn-
zeichnet, daß man Expressionskassetten enthaltend eine DNA-
Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder eine mit dieser
45 hybridisierende, bzw. zur Gesamtsequenz oder zu Teilsequenzen
homologen DNA-Sequenz in eine Pflanzenzelle oder Protoplasten

von Pflanzen einbringt und diese zu ganzen Pflanzen regeneriert.

- Verwendung der DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7
5 oder eine mit dieser hybridisierende DNA-Sequenz zur Herstellung von Pflanzen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen durch Expression einer MPMT DNA-Sequenz in Pflanzen.
- 10 Die Erfindung wird durch die nun folgenden Beispiele erläutert, ist aber nicht auf diese beschränkt:

Sequenzanalyse rekombinanter DNA

- 15 Die Sequenzierung rekombinanter DNA-Moleküle erfolgte mit einem Laserfluoreszenz-DNA-Sequenzierer der Firma Licor (Vertrieb durch MWG Biotech, Ebersbach) nach der Methode von Sanger (Sanger et al., Proc. Natl. Acad. Sci. USA 74 (1977), 5463 - 5467).

20 Beispiel 1

Identifizierung einer 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803.

- 25 Die Klonierung und Identifizierung der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803 erfolgte folgendermaßen:

- Unter Verwendung eines in S-Adenosyl-L-Methionin Methyltransferasen konservierten Sequenzmotivs, welches für die Bindung des
30 S-Adenosyl-L-Methionin (SAM) verantwortlich ist (C.P. Joshi und V.L. Chiang. PMB. 37: 663-374, 1998), wurde eine genomische DNA Datenbank von *Synechocystis spec.* PCC 6803 durchmustert (Kaneko et al., DNA Res. 34:109-136, 1996). Die bei der Durchmusterung
35 identifizierten hypothetischen Proteine, welche über das SAM-Bindemotiv verfügten, wurden mit den Primärsequenzen der *Synechocystis spec.* PCC 6803 γ -Tocopherol-methyltransferase (bezeichnet als slr0089) sowie der *Arabidopsis thaliana* γ -Tocopherol-methyltransferase (David Shintani und Dean DellaPenna. Science.
40 282:2098-2100, 1998) verglichen.

- Dabei konnte ein hypothetisches Protein identifiziert werden (bezeichnet sl10418 SEQ.-ID Nr. 2), welches geringe Übereinstimmung in der Aminosäuresequenz mit den γ -Tocopherol-methyltransferasen
45 aus *Synechocystis spec.* PCC 6803 und *Arabidopsis thaliana* aufwies (36% bzw. 28% Identität).

- Weitere Untersuchungen der Primärsequenz des hypothetischen Proteins sll0418 belegten das Vorkommen einer putativen prokaryontischen Signalsequenz innerhalb der ersten 20 Aminosäuren (PSIGNAL, PC/GENE™ IntelliGenetics, Inc ©1991). Eine solche Sequenz konnte
5 ebenfalls in der *Synechocystis spec. PCC 6803* γ -Tocopherolmethyltransferase (slr0089) identifiziert werden (D. Shintani und D. DellaPenna. *Science*. 282:2098-2100,1998) und deutet auf eine identische Lokalisation der beiden Proteine hin.
- 10 Das vorhergesagte Molekulargewicht des unprozessierten Proteins beträgt 34,9 kDa und liegt damit in einem Bereich der auch für die *Synechocystis spec. PCC 6803* γ -Tocopherolmethyltransferase (David Shintani und Dean DellaPenna, *Science*. 282:2098-2100,1998) und der aus Paprikafrüchten gereinigten γ -Tocopherolmethyltrans-
15 ferase (d'Harlingue and Camara, *Plastid enzymes of terpenoid biosynthesis: Purification of γ -Tocopherol Methyltransferase from Capsicum Chromoplasts. Journal of Biological Chemistry, Vol. 269 No.28, 15200-152003,1985)* ermittelt wurde.
- 20 Unter Berücksichtigung der Fakten, schlußfolgerten wir, daß es sich bei dem hypothetischen Protein sll0418 um eine Tocopherolmethyltransferase handeln könnte.

Beispiel 2

- 25 Amplifikation und Klonierung der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec. PCC 6803*

- Die DNA kodierend für den ORF (open reading frame) sll0418 wurde
30 mittels polymerase chain reaction (PCR) aus *Synechocystis spec. PCC 6803* gemäß der Methode nach Crispin A. Howitt (*BioTechniques* 21:32-34, July 1996) unter Verwendung eines sense spezifischen Primers (sll04185' Seq. Nr. 5) und eines antisense spezifischen Primers (sll04183' Seq. Nr. 6) amplifiziert.

- 35 Die PCR Bedingungen waren die folgenden:

Die PCR erfolgte in einem 50 μ l Reaktionsansatz in dem enthalten war:

- 40 -5 μ l einer *Synechocystis spec. PCC 6803* Zellsuspension
-0,2 mM dATP, dTTP, dGTP, dCTP
-1,5 mM Mg(OAc)₂
-5 μ g Rinderserum-Albumin
45 -40pmol sll04185'
-40pmol sll04183'
-15 μ l 3,3 \times rTth DNA Polymerase XLPuffer (PE Applied Biosystems)

-5U rTth DNA Polymerase XL (PE Applied Biosystems)

Die PCR wurde unter folgenden Zyklusbedingungen durchgeführt:

Schritt 1: 5 Minuten 94°C (Denaturierung)

Schritt 2: 3 Sekunden 94°C

5 Schritt 3: 2 Minuten 58°C (Annealing)

Schritt 4: 2 Minuten 72°C (Elongation)

40 Wiederholungen der Schritte 2-4

Schritt 5: 10 Minuten 72°C (Post-Elongation)

Schritt 6: 4°C (Warteschleife)

10

Das Amplikon wurde unter Verwendung von Standardmethoden in den PCR Klonierungsvektor pGEM-T (Promega) kloniert. Die Identität des erzeugten Amplikons wurde durch Sequenzierung unter Verwendung des M13F (-40) Primers bestätigt.

15

Beispiel 3

Erzeugung einer sll0418 Knock out Mutante

20 Ein DNA Konstrukt zur Erzeugung einer Deletionsmutante des ORF sll0418 in *Synechocystis spec.* PCC 6803 wurde unter Anwendung von Standard Klonierungstechniken erzeugt.

Der Vektor pGEM-T/sll0418 wurde unter Verwendung des Restriktions-
25 onsenzyms BallI verdaut. Das Vorhandensein von zwei BallI Schnittstellen innerhalb der sll0418 Sequenz (Position Bp 109 bzw Bp 202) hatte den Verlust eines 93 Bp umfassenden internen Fragmentes zur Folge. In die BallI Schnittstellen des sll0418 ORF wurde die Aminoglycosid-3'Phosphotransferase des Transposons Tn903 kloniert.
30 Dazu wurde das Tn903 als EcoRI Fragment aus dem Vektor pUC4k (Vieira, J und Messing, J Gene:19, 259-268, 1982) isoliert, die überstehenden Enden des Restriktionsverdaus nach Standardmethoden in glatte Enden überführt und in den BallI geschnittenen Vektor pGEM-T/sll0418 ligiert. Der Ligationsansatz wurde zur
35 Transformation von *E.coli* Xl1 blue Zellen verwendet. Transformanten wurden durch Verwendung von Kanamycin und Ampicillin selektioniert. Ein rekombinantes Plasmid (pGEM-T/sll0418::tn903) wurde isoliert und zur Transformation von *Synechocystis spec.* PCC 6803 gemäß der Methode nach Williams (Methods Enzymol. 167:776-778,
40 1987) eingesetzt.

Synechocystis spec. PCC 6803 Transformanten wurden selektioniert auf Kanamycin haltigem (kan) BG-11 Festmedium (Castenholz, Methods in Enzymology, Seite 68-93, 1988) bei 28°C und 30µmol
45 Photonen $\times (\text{m}^2 \times \text{s})^{-1}$. Vier unabhängige Knock out Mutanten konnten

nach fünf Selektionsrunden (Passagen von Einzelkolonien auf frisches BG-11kn Medium) erzeugt werden.

Der vollständige Verlust des sll0418 Endogens bzw. der Austausch
5 gegen die rekombinante sll0418::tn903 DNA, wurde durch PCR Analysen bestätigt.

Beispiel 4

10 Vergleich der Tocopherolproduktion in *Synechocystis spec.* PCC 6803 Wildtypzellen und den erzeugten Knock out Mutanten des ORF sll0418.

Die auf den BG-11kan Agarmedium kultivierten Zellen der vier un-
15 abhängigen *Synechocystis spec.* PCC 6803 Knock out Mutanten des ORF sll0418 sowie untransformierte Wildtypzellen wurden zum Animpfen von Flüssigkulturen verwendet. Diese Kulturen wurden bei 28°C und 30µmol Photonen $\times (m^2 \times s)^{-1}$ (30µE) für ca. 3 Tage kultiviert. Nach Bestimmung der OD₇₃₀ der einzelnen Kulturen, wurde die
20 OD₇₃₀ aller Kulturen durch entsprechende Verdünnungen mit BG-11 (Wildtypen) bzw. BG-11kan (Mutanten) synchronisiert. Diese auf Zelldichte synchronisierten Kulturen wurden zum Animpfen von drei Kulturen pro Mutante bzw. der Wildtypkontrollen verwendet. Die biochemischen Analysen konnten somit unter Verwendung von jeweils
25 drei unabhängig gewachsenen Kulturen einer Mutante und der entsprechenden Wildtypen durchgeführt werden. Die Kulturen wurden bis zu einer optischen Dichte von OD₇₃₀=0,3 angezogen. Das Medium der Zellkultur wurde durch zweimalige Zentrifugation bei 14000 rpm in einer Eppendorf Tischzentrifuge entfernt. Der daran
30 anschließende Aufschluß der Zellen erfolgte durch viermalige Inkubation im Eppendorfschüttler bei 30°C, 1000rpm in 100% Methanol für 15 Minuten, wobei die jeweils erhaltenen Überstände vereinigt wurden. Weitere Inkubationsschritte ergaben keine weitere Freisetzung von Tocopherolen oder Tocotrienolen.

35 Um Oxidation zu vermeiden, wurden die erhaltenen Extrakte direkt nach der Extraktion mit Hilfe einer Waters Alliance 2690 HPLC-Anlage analysiert. Tocopherole und Tocotrienole wurden über eine reverse Phase Säule (ProntoSil 200-3-C30, Bischoff) mit einer mo-
40 bilen Phase von 100% Methanol getrennt und anhand von Standards (Merck) identifiziert. Als Detektionssystem diente die Fluoreszenz der Substanzen (Anregung 295nm, Emission 320 nm), die mit Hilfe eines Jasco Fluoreszenzdetektors FP 920 nachgewiesen wurde.

In den *Synechocystis spec.* PCC 6803 knock out Mutanten des ORF sll0418 konnten keine Tocopherole und Tocotrienole gefunden werden. Tocopherole und Tocotrienole wurden jedoch in den *Synechocystis spec.* PCC 6803 Wildtypzellen gemessen.

5

Der Verlust der Fähigkeit zur Produktion von Tocopherolen und Tocotrienolen innerhalb der knock out Mutanten des ORF sll0418 im Vergleich zu den *Synechocystis spec.* PCC 6803 Wildtypzellen zeigt, daß das Gen sll0418 für eine 2-Methyl-6-phytylhydrochinon-

10 methyltransferase kodiert.

Beispiel 5

Funktionelle Charakterisierung der 2-Methyl-6-phytylhydrochinon-
15 methyltransferase aus *Synechocystis spec.* PCC 6803 durch heterologe Expression in *E.coli*.

Das hypothetische Protein sll0418 aus *Synechocystis spec.* PCC 6803 konnte durch funktionelle Expression in *E.coli* als
20 2-Methyl-6-phytylhydrochinon-methyltransferase identifiziert werden.

Das aus *Synechocystis spec.* PCC 6803 amplifizierte Gen sll0418 wurde im korrekten Leserahmen in den Expressionsvektor pQE-30
25 (Qiagen) subkloniert. Die zur Amplifikation des ORF sll0418 aus *Synechocystis spec.* PCC 6803 verwendeten Primer sll04185' bzw. sll04183' (Sequenz ID Nr. 5 und 6) waren so konstruiert, daß an das 5' Ende und das 3' Ende des Amplikons BamHI Restriktions-
schnittstellen addiert wurden, siehe Sequenz ID Nr. 3. Das
30 sll0418 Fragment wurde unter Verwendung dieser flankierenden BamHI Restriktionschnittstellen aus dem rekombinanten Plasmid pGEM-T/sll0418 isoliert und unter Anwendung von Standardmethoden in einen BamHI geschnittenen pQE-30 ligiert. Der Ligationsansatz wurde zur Transformation von M15 *E.coli* Zellen verwendet und
35 Kanamycin und Ampicillin resistente Transformanten wurden analysiert. Die Kanamycin Resistenz wird durch das in den M15 Zellen enthaltene pREP-4 Plasmid vermittelt. Ein rekombinantes Plasmid (pQE-30/sll0418) welches das sll0418 Fragment in der richtigen Orientierung trug, wurde isoliert. Die Identität und Orientie-
40 rung des Inserts wurde durch Sequenzierung bestätigt.

Das rekombinante Plasmid pQE-30/sll0418 wurde zur Transformation von M15 *E.coli* Zellen verwendet, um rekombinantes sll0418 Protein zu erzeugen. Unter Verwendung einer aus der Transformation
45 hervorgegangenen Kolonie wurde eine Übernachtskultur in Luria Broth Medium mit 200µg/ml Ampicillin (Amp) und 50µg/ml Kanamycin (Kan) angeimpft. Ausgehend von dieser Kultur wurde am nächsten

Morgen eine 100ml Luria Broth Kultur (Amp/Kan) angeimpft. Diese Kultur wurde bei 28°C auf einem Schüttelinkubator bis zum Erreichen einer OD₆₀₀:0,35-0,4 inkubiert. Anschließend wurde die Produktion des rekombinanten Proteins durch Zugabe von 0,4 mM Iso-
5 propyl-β-D-thiogalaktopyranosid (IPTG) induziert. Die Kultur wurde für weitere 3 Stunden bei 28°C geschüttelt und die Zellen anschließend durch Zentrifugation bei 8000g pelletiert.

Das Pellet wurde in 600µl Lysispuffer (ca. 1-1,5 ml /g Pellet
10 Naßgewicht, 10 mM HEPES KOH pH 7,8, 5 mM Dithiothreitol (DTT), 0,24 M Sorbitol) resuspendiert. Anschließend wurde PMSF (Phenylmethylsulfonat) zu einer Endkonzentration von 0,15 mM beige-
fügt und der Ansatz für 10 Minuten auf Eis gestellt. Der Auf-
schluß der Zellen erfolgte durch einen 10 Sekunden Ultraschall-
15 Puls unter Verwendung eines Ultraschallstabes. Nach Zugabe von Triton X100 (Endkonzentration 0,1%) wurde die Zellsuspension für 30 Minuten auf Eis inkubiert. Der Ansatz wurde anschließend für 30 Minuten bei 25000xg abzentrifugiert und der Überstand zum Assay eingesetzt.

20

Die Aktivitätsbestimmung der 2-Methyl-6-phytylhydrochinon-methyltransferase erfolgt durch Nachweis des radioaktiv markierten Reaktionsproduktes 2,3-Dimethyl-6-phytylhydrochinon.

25 Dazu wurden 135µl des Enzyms (ca.300-600µg) zusammen mit 20µl Substrat (2-Methyl-6-phytylhydrochinon) und 15µl (0,46 mM SAM ¹⁴C) Methylgruppendonor in folgendem Reaktionspuffer : 200µl (125mM) Tricine-NaOH pH 7,6, 100µl (1,25 mM) Sorbitol, 10µl (50mM) MgCl₂ und 20µl (250mM) Ascorbat für 4 Stunden bei 25°C im Dunkeln
30 inkubiert.

Das Abstoppen der Reaktion erfolgte durch Zugabe von 750µl Chloroform/Methanol (1:2) + 150µl 0,9% NaCl. Der gemischte Ansatz wurde kurz zentrifugiert und die obere Phase wurde verworfen. Die
35 untere Phase wird in ein neues Reaktionsgefäß überführt und unter Stickstoff eingedampft. Die Rückstände wurden in 20µl Ether aufgenommen und auf eine Dünnschicht-Platte zur chromatographischen Trennung der Substanzen aufgetragen (feste Phase: HPTLC-Platten: Kieselgel 60 F₂₅₄ (Merk), flüssige Phase: Toluol). Der Nachweis
40 des radioaktiv markierten Reaktionsproduktes erfolgt durch Verwendung eines Phosphorimagers.

Diese Experimente bestätigten, daß es sich bei dem durch das Gen sl10418 (SEQ-ID Nr.1) aus *Synechocystis spec.* PCC 6803 kodierte
45 Protein um eine 2-Methyl-6-phytylhydrochinon-methyltransferase handelt, da es die enzymatische Aktivität zur Umwandlung von

2-Methyl-6-phytylhydrochinon in 2,3-Dimethyl-6-phytylhydrochinon besitzt.

Abbildung 2 zeigt einen Sequenzvergleich auf Aminosäureebene zwischen den γ -Tocopherolmethyltransferasen aus *Synechocystis spec.* PCC *Synechocystis spec.* PCC 6803 (slr0089) und *A.thaliana* (aratmt) mit der 2-Methyl-6-phytylhydrochinon-methyltransferase (sl104189) aus *Synechocystis spec.* PCC 6803. Die Übereinstimmung mit den γ -Tocopherolmethyltransferasen aus *Synechocystis spec.* PCC 6803 und *Arabisopsis thaliana* beträgt 36 bzw. 28 % Identität.

Beispiel 6

Substratspezifität der 2-Methyl-6-phytylhydrochinon-methyltransferase

Enzymatische Untersuchungen wie in Beispiel 5 durchgeführt belegen, daß das Enzym MPMT - kodiert durch das Gen sl10418 (SEQ-ID Nr. 1) aus *Synechocystis spec.* PCC 6803 - 2-Methyl-6-phytylhydrochinon in 2,3-Dimethyl-6-phytylhydrochinon umwandelt.

Zusätzlich besitzt das Enzym MPMT eine 2-Methyl-6-geranylgeranylhydrochinon-methyltransferase Aktivität, wohingegen eine γ -Tocopherolmethyltransferase Aktivität nicht nachgewiesen werden konnte. Somit ist belegt, daß das Enzym 2-Methyl-6-phytylhydrochinon-methyltransferase an der Biosynthese der Tocotrienole beteiligt ist, da es 2-Methyl-6-geranylgeranylhydrochinon zu 2,3-Dimethyl-6-geranylgeranylhydrochinon umwandelt. Dies zeigt deutlich die Verschiedenheit der Enzymaktivität der 2-Methyl-6-phytylhydrochinon-methyltransferase im Vergleich zur γ -Tocopherolmethyltransferase.

Beispiel 7

Herstellung von Expressionskassetten enthaltend das MPMT-Gen

Transgene Pflanzen wurden erzeugt, die die 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC6803 zum einen unter Kontrolle des konstitutiven 35S-Promotor des CaMV (Blumenkohlmosaikvirus) (Franck et al., Cell 21: 285-294, 1980) und zum anderen unter Kontrolle des samenspezifischen Promotors des Legumin Gens aus *Vicia faba* (Kafatos et al., Nuc. Acid. Res., 14(6): 2707-2720, 1986) exprimieren. Die Grundlage der konstitutiven Expression der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803 erzeugten Plasmides war der pBinAR-TkTp-9 (Ralf Badur, Dissertation Universität Göttingen, 1998). Dieser Vektor ist ein Derivat des pBinAR (Höfgen

und Willmitzer, Plant Sci. 66: 221-230, 1990) und enthält den 35S-Promotor des CaMV (Blumenkohlmosaikvirus) (Franck et al., 1980), das Terminationssignal des Octopin-Synthase Gens (Gielen et al., EMBO J. 3: 835-846, 1984) sowie die für das Transitpeptid der plastidären *Nicotiana tabacum* Transketolase kodierende DNA Sequenz (Ralf Badur, Dissertation Universität Göttingen, 1998). Die unter Berücksichtigung des korrekten Leserasters erfolgte Klonierung der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC6803 in diesen Vektor, erzeugt eine Translationsfusion der 2-Methyl-6-phytylhydrochinon-methyltransferase mit dem plastidären Transitpeptid. Dadurch erfolgt ein Transport des Transgens in die Plastiden.

Zur Erstellung dieses Plasmides wurde das Gen sll0418 unter Verwendung der flankierenden BamHI Restriktionsschnittstellen aus dem Plasmid pGEM-T/sll0418 isoliert. Dieses Fragment wurde unter Anwendung von Standardmethoden in einen BamHI geschnittenen pBinAR-TkTp-9 ligiert (siehe Abbildung 3). Dieses Plasmid (pBinAR-TkTp-9/sll0418) wurde zur Erzeugung transgener *Arabidopsis thaliana*, *Brassica napus* und *Nicotiana tabacum* verwendet. Fragment A (529 bp) in Abbildung 3 beinhaltet den 35S-Promotor des CaMV (Nukleotide 6909 bis 7437 des Blumenkohlmosaikvirus), Fragment B (245bp) kodiert für das Transitpeptid der *Nicotiana tabacum* Transketolase, Fragment C (977Bp) kodiert ORF sll0418 aus *Synechocystis spec.* PCC 6803, Fragment D (219Bp) kodiert für das Terminationssignal des Octopin-Synthase Gens.

Zur Erzeugung eines Plasmides, welches die samenspezifische Expression der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803 in Pflanzen ermöglicht, wurde der samenspezifische Promotor des Legumin B4 Gens (Kafatos et al., Nuc. Acid. Res., 14(6):2707-2720, 1986) verwendet. Aus dem Plasmid pCR-Script/lePOCS wurde das 2,7 Kb Fragment des Legumin B4 Gen Promotors unter Verwendung der den Promotor 5' flankierenden EcoRI und der 3' flankierenden KpnI Schnittstellen isoliert. Das Plasmid pBinAR-TkTp-9/sll0418 wurde ebenfalls mit den Restriktionsenzymen EcoRI und KpnI behandelt. Dies hatte zur Folge, daß der 35S-Promotor des CaMV aus diesem Plasmid herausgetrennt wurde. Der Promotor des Legumin Gens wurde anschließend als EcoRI/KpnI Fragment in diesen Vektor kloniert, wodurch ein Plasmid erzeugt wurde, welches die Expression des Gen sll0418 unter die Kontrolle dieses samenspezifischen Promotors stellte, siehe Abbildung 4. Dieses Plasmid (pBinARleP-TkTp-9/sll0418) wurde zur Erzeugung transgener *Arabidopsis thaliana*, *Brassica napus* und *Nicotiana tabacum* Pflanzen verwendet.

Fragment A (2700 bp) in Abbildung 4 beinhaltet den Promotor des Legumin B4 Gens aus *Vicia faba*, Fragment B (245bp) kodiert für das Transitpeptid der *Nicotina tabacum* Transketolase, Fragment C (977Bp) kodiert für das ORF sll0418 aus *Synechocystis spec. PCC 6803*, Fragment D (219Bp) für das Terminationssignal des Octopin-Synthase Gens.

Beispiel 8

10 Herstellung von Expressionskassetten enthaltend einen Deletionsklon des MPMT-Gens

Auf Grundlage einer Computeranalyse wurde in der Primärsequenz des ORF sll0418 ein putatives prokaryontisches Sekretionssignal identifiziert. Um sicherzustellen, daß dieses bei der Expression in Pflanzen keinen negativen Einfluß auf den Import des Proteins in die Plastiden nimmt, wurde ein Derivat der Sequenz des sll0418 erzeugt, bei dem das putative Sekretionssignal deletiert wurde (Sequenz-ID Nr. 7). Diese Deletion wurde unter Anwendung der PCR Technologie durchgeführt. Durch die dabei verwendeten Primer (sll0418Dsp5', Sequenz-ID Nr. 9 und sll0418Dsp3', Sequenz-ID Nr. 10) wurde an das 5'Ende der Sequenz eine EcoRV Restriktionsschnittstelle und an das 3'Ende eine SalI Restriktionsschnittstelle addiert, durch die eine gerichtete Klonierung in den Vektor pBinAR-TkTp-9 ermöglicht wurde. Das entstandene Plasmid pBinAR-TkTp-9/sll0418ASP ist in Abbildung 5 beschrieben. Fragment A (529 bp) in Abbildung 5 beinhaltet den 35S-Promotor des CaMV (Nukleotide 6909 bis 7437 des Blumenkohlmosaikvirus), Fragment B (245bp) Fragment kodiert für das Transitpeptid der *Nicotiana tabacum* Transketolase, Fragment C (930Bp) ORF sll0418ASP aus *Synechocystis spec. PCC 6803* Fragment D (219Bp) für das Terminationssignal des Octopin-Synthase Gens.

Zur Erzeugung eines Plasmides, welches die samenspezifische Expression des Deletionsklons der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec. PCC6803* in Pflanzen ermöglicht, wurde ebenfalls der bereits beschriebene samenspezifische Promotor des Legumin B4 Gens (Kafatos et al., Nuc. Acid. Res., 14(6):2707-2720, 1986) verwendet. Aus dem Plasmid PCR-Script/lePOCS wurde das 2,7 Kb Fragment des Legumin B4 Gen Promotors unter Verwendung der den Promotor 5' flankierenden EcoR1 und der 3' flankierenden Kpn1 Schnittstellen isoliert. Das Plasmid pBinAR-TkTp-9/sll0418ASP wurde ebenfalls mit den Restriktionsenzymen EcoR1 und Kpn1 behandelt. Dies hatte zur Folge, daß der 35S-Promotor des CaMV aus diesem Plasmid herausgetrennt wurde. Der Promotor des Legumin Gens wurde anschließend als EcoR1/Kpn1 Fragment in diesen Vektor kloniert, wodurch ein

Plasmid erzeugt wurde, welches die Expression des Deletionsklons des Gen sll0418 unter die Kontrolle dieses samenspezifischen Promotors stellte, siehe Abbildung 6. Fragment A (2700 bp) in Abbildung 6 beinhaltet den Promotor des Legumin B4 Gens aus *Vicia faba*, Fragment B (245bp) Fragment kodiert für das Transitpeptid der *Nicotiana tabacum* Transketolase, Fragment C (930Bp) ORF sll0418ASP aus *Synechocystis spec.* PCC 6803 Fragment D (219Bp) für das Terminationssignal des Octopin-Synthase Gens.

- 10 Dieses Plasmid (pBinARleP-TkTp-9/sll0418ASP) wurde zur Erzeugung transgener *Arabidopsis thaliana*, *Brassica napus* und *Nicotiana tabacum* Pflanzen verwendet.

Auch durch Expression der DNA-Sequenz SEQ-ID Nr. 7 in transgenen Pflanzen wurde eine Steigerung des Gehaltes an Tocopherol und Tocotrienol gemessen.

Beispiel 9

- 20 Herstellung transgener *Arabidopsis thaliana* Pflanzen

Wildtyp *Arabidopsis thaliana* Pflanzen (Columbia) wurden mit dem *Agrobacterium tumefaciens* Stamm (EHA105) auf Grundlage einer modifizierten Vacuum Infiltrationsmethode transformiert (Steve Clough und Andrew Bent, Floral dip: a simplified method for *Agrobacterium* mediated transformation of *Arabidopsis thaliana*. Plant J. 16(6):735-43, 1998; Bechtold, N., Ellis, J. und Pelltier, G., in: *Planta Agrobacterium-mediated gene transfer by infiltration of adult Arabidopsis thaliana plants*. CR Acad Sci Paris, 1993. 1144(2):204-212). Die verwendeten *Agrobacterium tumefaciens* Zellen waren im Vorfeld mit den Plasmiden pBinARleP-TkTp-9/sll0418 bzw. pBinAR-TkTp-9/sll0418 (Abbildung 3 und 4) transformiert worden.

- 35 Samen der Primärtransformanten wurden auf Grundlage der Antibiotikaresistenz selektioniert. Antibiotika resistente Keimlinge wurden in Erde gepflanzt und als vollentwickelte Pflanzen zur biochemischen Analyse verwendet.

- 40 Beispiel 10

Herstellung transgener *Brassica napus* Pflanzen

Die Herstellung transgener Raps Pflanzen orientierte sich an einem Protokoll von Bade, J.B. und Damm, B. (in *Gene Transfer to Plants*, Potrykus, I. und Spangenberg, G., eds, Springer Lab Ma-

nual, Springer Verlag, 1995, 30-38), in welchem auch die Zusammensetzung der verwendeten Medien und Puffer angegeben ist.

Die Transformationen erfolgten mit dem *Agrobacterium tumefaciens* Stamm EHA105. Zur Transformation wurden die Plasmide pBinARleP-TkTp-9/sll0418 bzw. pBinAR-TkTp-9/sll0418 verwendet. Samen von *Brassica napus* var. Westar wurden mit 70% Ethanol (v/v) oberflächensteril gemacht, 10 Minuten bei 55°C in Wasser gewaschen, in 1%iger Hypochlorit-Lösung (25% v/v Teepol, 0,1% v/v Tween 20) für 20 Minuten inkubiert und sechsmal mit sterilem Wasser für jeweils 20 Minuten gewaschen. Die Samen wurden drei Tage auf Filterpapier getrocknet und 10-15 Samen in einem Glaskolben mit 15 ml Keimungsmedium zur Keimung gebracht. Von mehreren Keimlingen (ca. 10 cm groß) wurden die Wurzeln und Apices entfernt und die verbleibenden Hypokotyle in ca. 6 mm lange Stücke geschnitten. Die so gewonnenen ca. 600 Explantate wurden 30 Minuten mit 50 ml Basalmedium gewaschen und in einen 300 ml Kolben überführt. Nach Zugabe von 100 ml Kallusinduktionsmedium wurden die Kulturen für 24 Stunden bei 100 U/min inkubiert.

Vom *Agrobacterium* Stamm wurde eine Übernachtskultur bei 29°C in Luria Broth-Medium mit Kanamycin (20mg/l) angesetzt, davon 2 ml in 50 ml Luria Broth-Medium ohne Kanamycin für 4 Stunden bei 29°C bis zu einer OD₆₀₀ von 0,4-0,5 inkubiert. Nach der Pelletierung der Kultur bei 2000 U/min für 25 min wurde das Zellpellet in 25 ml Basalmedium resuspendiert. Die Konzentration der Bakterien in der Lösung wurde durch Zugabe von weiterem Basalmedium auf eine OD₆₀₀ von 0,3 eingestellt.

Aus den Raps-Explanten wurde das Kallus-Induktionsmedium mit sterilen Pipetten entfernt, 50 ml *Agrobacterium*-Lösung hinzugefügt, vorsichtig gemischt und für 20 min inkubiert. Die *Agrobacterien*-Suspension wurde entfernt, die Raps-Explante für 1 min mit 50 ml Kallus-Induktionsmedium gewaschen und anschließend 100 ml Kallus-Induktionsmedium hinzugefügt. Die Co-Kultivierung wurde für 24 h auf einem Rotationsschüttler bei 100 U/min durchgeführt. Die Co-Kultivierung wurde durch Wegnahme des Kallus-Induktionsmediums gestoppt und die Explante zweimal für jeweils 1 min mit 25 ml und zweimal für 60 min mit jeweils 100 ml Waschmedium bei 100 U/min gewaschen. Das Waschmedium mit den Explanten wurde in 15 cm Petrischalen überführt und das Medium mit sterilen Pipetten entfernt.

Zur Regeneration wurden jeweils 20-30 Explante in 90 mm Petrischalen überführt, welche 25 ml Sproß-Induktionsmedium mit Kanamycin enthielten. Die Petrischalen wurden mit 2 Lagen Leukopor verschlossen und bei 25 °C und 2000 lux bei Photoperioden von 16

Stunden Licht/ 8 Stunden Dunkelheit inkubiert. Alle 12 Tage wurden die sich entwickelnden Kalli auf frische Petrischalen mit Sproß-Induktionsmedium umgesetzt. Alle weiteren Schritte zur Regeneration ganzer Pflanzen wurden wie von Bade, J.B und Damm, B. (in: Gene Transfer to Plants, Potrykus, I. und Spangenberg, G., eds, Springer Lab Manual, Springer Verlag, 1995, 30-38) beschrieben durchgeführt.

Beispiel 11

10

Herstellung transgener *Nicotiana tabacum* Pflanzen

Zehn ml YEB-Medium mit Antibiotikum (5 g/l Rinder-Extrakt, 1 g/l Hefe-Extrakt, 5 g/l Pepton, 5 g/l Saccharose und 2 mM $MgSO_4$) wurden mit einer Kolonie von *Agrobacterium tumefaciens* beimpft und über Nacht bei 28°C kultiviert. Die Zellen wurden 20 min bei 4°C, 3500 U/min in einer Tischzentrifuge pelletiert und danach in frischem YEB-Medium ohne Antibiotika unter sterilen Bedingungen resuspendiert. Die Zellsuspension wurde für die Transformation eingesetzt.

Die Wildtyp-Pflanzen aus Sterilkultur wurden durch vegetative Replikation erhalten. Dazu wurde nur die Spitze der Pflanze abgeschnitten und auf frisches 2MS-Medium in ein steriles Einweckglas überführt. Vom Rest der Pflanze wurden die Haare auf der Blattoberseite und die Mittelrippen der Blätter entfernt. Die Blätter wurden mit einer Rasierklinge in etwa 1 cm² große Stücke geschnitten. Die Agrobakterienkultur wurde in eine kleine Petrischale überführt (Durchmesser 2 cm). Die Blattstücke wurden kurz durch die Lösung gezogen und mit der Blattunterseite auf 2MS-Medium in Petrischalen (Durchmesser 9 cm) gelegt, so daß sie das Medium berührten. Nach zwei Tagen im Dunkeln bei 25°C wurden die Explantate auf Platten mit Kallusinduktionsmedium überführt und in der Klimakammer auf 28°C temperiert. Das Medium mußte alle 7-10 Tage gewechselt werden. Sobald sich Kalli bildeten, wurden die Explantate in sterile Einweckgläser auf Sproßinduktionsmedium mit Claforan (siehe oben) überführt. Nach etwa einem Monat trat Organogenese ein und die gebildeten Sprossen konnten abgeschnitten werden. Die Kultivierung der Sprosse wurde auf 2MS-Medium mit Claforan und Selektionsmarker durchgeführt. Sobald sich ein kräftiger Wurzelballen gebildet hatte, konnte die Pflanzen in Pikiererde getopft werden.

Beispiel 12

45

Charakterisierung der transgenen Pflanzen

Um zu bestätigen, daß durch die Expression der 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803 die Vitamin E Biosynthese in den transgenen Pflanzen gesteigert wird, wurden die Tocopherol- und Tocotrienol-Gehalte in Blätter und Samen der mit den Konstrukten pBinARleP-TkTp-9/sll0418 bzw. pBinAR-TkTp-9/sll0418 Pflanzen (*Arabidopsis thaliana*, *Brassica napus* und *Nicotiana tabacum*) analysiert. Dazu wurden die transgenen Pflanzen im Gewächshaus kultiviert und Pflanzen die das Gen kodierend für die 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis spec.* PCC 6803 exprimieren auf Northern-Ebene analysiert. In Blättern und Samen dieser Pflanzen wurde der Tocopherolgehalt und der Tocotrienolgehalt ermittelt. In allen Fällen war die Tocopherol- bzw. Tocotrienol-Konzentration in transgenen Pflanzen, die zusätzlich eine DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 exprimieren, im Vergleich zu nicht transformierten Pflanzen erhöht.

20

25

30

35

40

45

Patentansprüche

1. DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 und mit dieser hybridisierende oder zur Gesamtsequenz oder zu Teilsequenzen homologen DNA-Sequenz kodierend für eine 2-Methyl-6-phytylhydrochinon-methyltransferase aus *Synechocystis*.
5
2. Verwendung von DNA-Sequenzen codierend für eine 2-Methyl-6-phytylhydrochinon-methyltransferase zur Herstellung von Pflanzen und photosynthetisch aktiven Organismen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen.
10
3. Verwendung einer DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder einer mit dieser hybridisierenden DNA-Sequenz kodierend für eine 2-Methyl-6-phytylhydrochinon-methyltransferase zur Herstellung von Pflanzen und photosynthetisch aktiven Organismen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen.
15
4. Verfahren zur Herstellung von Pflanzen und photosynthetisch aktiven Organismen mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen dadurch gekennzeichnet, daß eine DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder eine mit dieser hybridisierende oder zur Gesamtsequenz oder zu Teilsequenzen homologen DNA-Sequenz in Pflanzen und photosynthetisch aktiven Organismen exprimiert wird.
20
25
5. Verfahren zur Transformation einer Pflanze dadurch gekennzeichnet, daß man eine Expressionskassette enthaltend einen Promotor, eine Signalsequenz, eine DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 und einen Terminator oder eine mit dieser hybridisierende DNA-Sequenz in eine Pflanzenzelle, in Kallusgewebe, eine ganze Pflanze oder Protoplasten von Pflanzenzellen einbringt.
30
35
6. Verfahren zur Transformation von Pflanzen gemäß Anspruch 5, dadurch gekennzeichnet, daß die Transformation mit Hilfe des Stammes *Agrobacterium tumefaciens*, der Elektroporation oder der particle bombardment Methode erfolgt.
40
7. Pflanze mit erhöhtem Gehalt an Tocopherolen und Tocotrienolen enthaltend eine Expressionskassette gemäß Anspruch 5.
45

Zeichn.

8. Pflanze nach Anspruch 7, ausgewählt aus der Gruppe Soja, Canola, Gerste, Hafer, Weizen, Raps, Mais, Roggen, Tagetes oder Sonnenblume.
- 5 9. Verwendung der DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder einer mit dieser hybridisierende DNA-Sequenz gemäß Anspruch 1 zur Herstellung eines Testsystems zur Identifizierung von Inhibitoren der 2-Methyl-6-phytylhydrochinon-methyltransferase.
- 10 10. Testsystem basierend auf der Expression der DNA-Sequenz SEQ-ID Nr. 1 oder SEQ-ID Nr. 7 oder einer mit dieser hybridisierende DNA-Sequenz gemäß Anspruch 1 zur Identifizierung von Inhibitoren der 2-Methyl-6-phytylhydrochinonmethyltransferase.
- 15

20

25

30

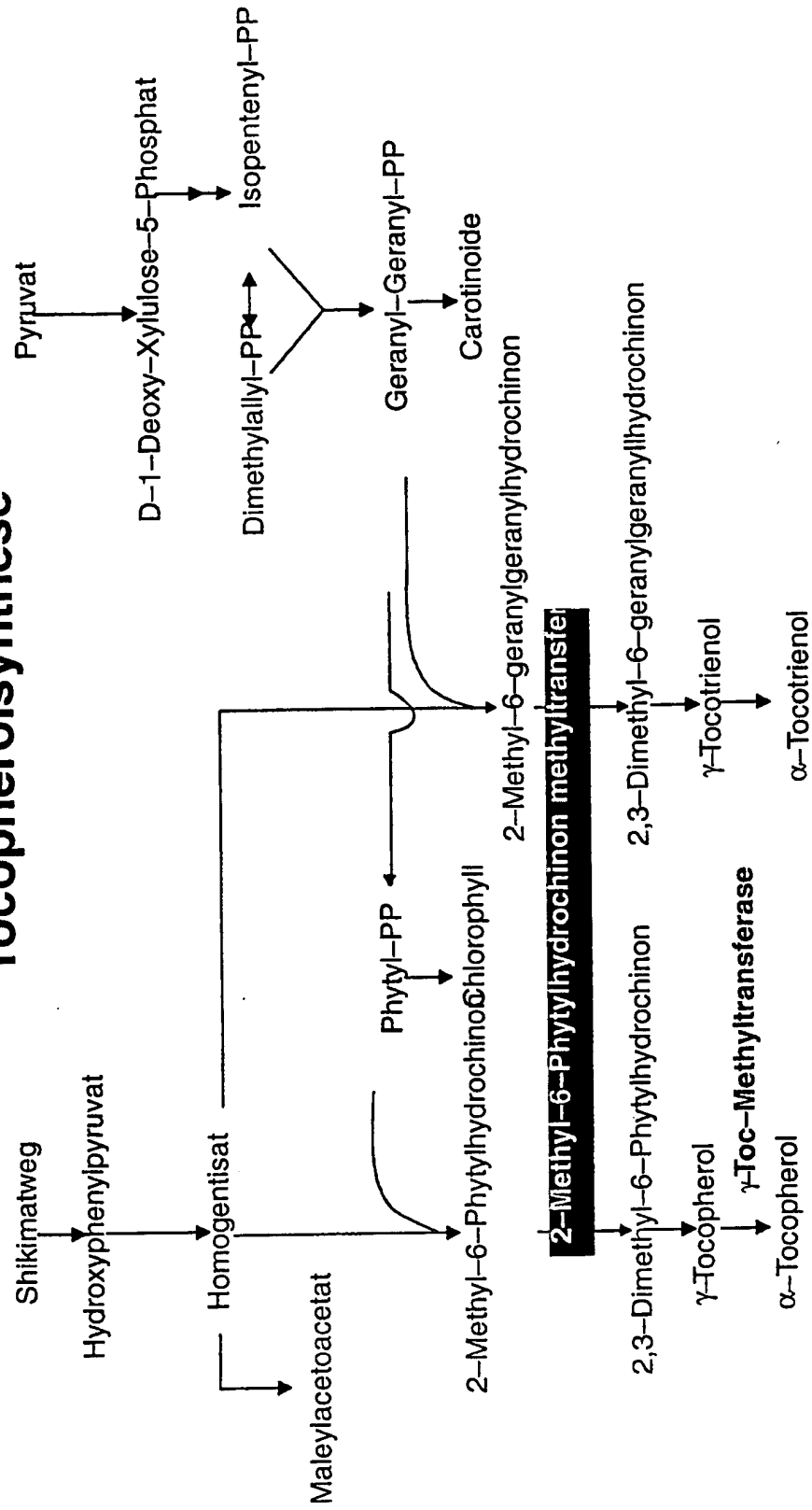
35

40

45

Tocopherolsynthese

Abbildung 1



Majority

[illegible][illegible]

Majority

X	X	X	X	X	X	X	X	L	X	X	A	X	A	X	T	X	X	X	X	X	X	X	X	X	X	X	S	X	G	X	X	E	X	X	W
								50						60																					
					</																														

14	L	S	L	A	I	A	A	G	L	Y	L	L	T	A	R	G	Y	Q	S	S	D	S	V	A	N	A	Y	D	Q	W	T	E	D	G	I	L	E	Y	W	sl10418.PRO	
14	A	F	Y	C	Y	F	S	L	L	T	M	A	S	A	T	I	A	S	A	D	L	Y	E	K	I	K	N	F	Y	D	D	S	S	G	L	W	E	D	V	W	slr0089.PRO
41	M	T	T	T	R	C	N	V	A	V	A	A	A	A	T	S	T	E	A	L	R	K	C	I	A	E	F	F	Y	N	E	T	S	G	L	W	E	I	W	G	aratnt.PRO

Majority

G	X	H	X	H	X	G	Y	G	D	X	X	V	X	X	D	-	-	X	X	A	Q	I	X	X	I	X	E	X	L	A	X	A	G	X	X	D
								90									100										110								120	

54	G	D	H	I	H	L	G	H	Y	G	D	P	P	V	A	K	-	-	-	D	F	I	Q	S	K	I	D	F	V	H	A	M	A	Q	W	G	L	D	sl10418.PRO		
54	G	E	H	M	H	H	G	Y	Y	G	P	H	G	T	Y	R	I	D	-	-	R	R	Q	A	Q	I	D	L	I	K	E	L	L	A	W	A	V	P	Q	N	slr0089.PRO
31	D	H	M	H	H	G	F	Y	D	P	D	S	S	V	Q	L	S	D	S	G	H	K	E	A	Q	I	R	M	I	E	E	S	L	R	F	A	G	V	T	D	aratmt.PRO

Majority

XX - - X K X X K V L D V G C G I G C S S R Y L A X X X G A E V X G I T L S P V

90	T	L	-	-	P	P	G	T	T	V	L	D	V	G	C	G	I	G	G	S	S	R	I	L	A	K	D	Y	G	F	N	V	T	G	I	T	I	S	P	Q	sl10418.PRO
92	S	A	-	-	-	K	P	R	K	I	L	D	L	G	C	G	I	G	G	S	S	L	Y	L	A	Q	Q	H	Q	A	E	V	M	G	A	S	L	S	P	V	slr0089.PRO
121	E	E	E	K	K	I	K	K	V	V	D	V	G	C	G	I	G	G	S	S	R	Y	L	A	S	K	F	G	A	E	C	I	G	I	T	L	S	P	V	aratmt.PRO	

Q V K R A X E L A X A X X L X X T A X F Q V A D A L D L P F X D G S F D X V W S Majority

170 180 190 200

128 Q V K R A T E L T P P - - D V T A K F A V D D A M A L S F P D G S F D V V W S sll0418.PRO
 129 Q V E R A C E R A R A L G L G S T C Q F Q V A N A L D L P F A S D S F D W V W S slr0089.PRO
 161 Q A K R A N D L A A A Q S L S H K A S F Q V A D A L D Q P F E D G K F D L V W S aratmt.PRO

X E S G E H M P D K A X F X K K E L X R V X K P G G R L I X A T W C H R X X X X G Majority

210 220 230 240

165 V E A G P H H P D K A V F A K K E L L R V V K P G G I L V V A D W N Q R D D R Q V sll0418.PRO
 169 L E S G E H M P N K A Q F L Q E A W R V L K P G G R L I L A T W C H R P I D P G slr0089.PRO
 201 M E S G E H M P D K A K F F V K E L V R V A A P G G R I I I V T W C H R N L S A G aratmt.PRO

X X X L X X X E X X X L X X X I X X X X X X L P A X X S X X X D Y X X X X X X X X X G Majority

250 260 270 280

205 P L N F W E K P V M R Q L L D Q W S H P A F A S I E G F A E N L E A T G L V E G sll0418.PRO
 209 N G P L T A D E R R H L Q A I Y D V Y C L P Y V V S L P D Y E A I A R E C G F G slr0089.PRO
 241 E E A L Q P W E Q N I L D K I C K T F Y L P A W C S T D D Y V N L L Q S H S L Q aratmt.PRO

X I K T A D W S V X X V A P F W X X V I X X A X X X X L W X L X X X G X K I I X Majority

290 300 310 320

245 Q V T T A D W T V P T L P A W L D T I W Q G I I R P Q G W L Q Y G I R G F I K S sll0418.PRO
 249 E I K T A D W S V A V A P F W D R V I E S A F D P R V L W A L G Q A G P K I I N slr0089.PRO
 281 D I K C A D W S E N V A P F W P A V I R T A L T W K G L V S L L R S G M K S I K aratmt.PRO

	X A L X X X L M X X G Y X X - - G L X R F - - G X X T X X K P L X X - X - - - -	Majority
	330 340 350 360	
285	V R E V P T I L L M R L A F G V G L C R F - - G M F K A V R K N A T Q A	sl10418.
289	A A L C L R L M K W G Y E R - - G L V R F - - G L L T G I K P L V - . S P Q S P	slr0089.
321	G A L T H P L M I E G Y K K - - G V I K F - - G I I T C Q K P L	aratmt.l

Majority

318		sl10418.
324	R	slr0089.
348		aratmt.l

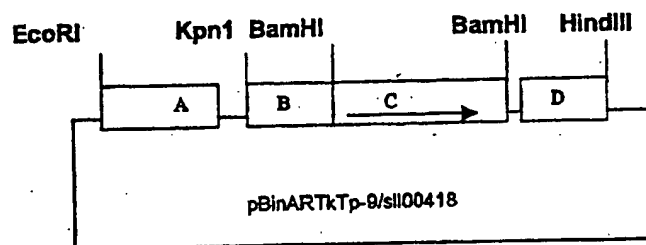


Abbildung 4

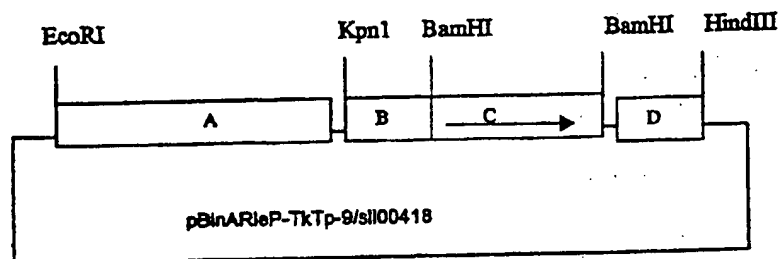


Abbildung 5

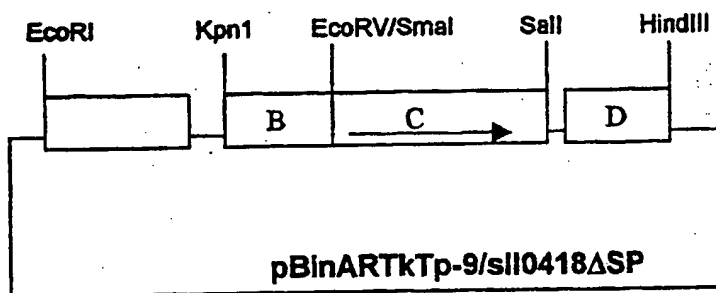
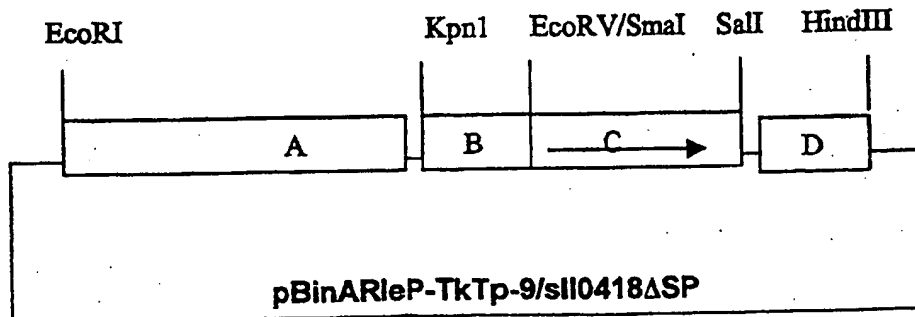


Abbildung 6



SEQUENZPROTOKOLL

<110> SunGene GmbH & Co.KGaA

<120> Ueberexpression einer DNA-Sequenz codierend fuer eine
2-Methyl-phytylhydrochinon-methyltransferase in
Pflanzen.

<130> MPMTSynechocystis

<140>

<141>

<160> 10

<170> PatentIn Vers. 2.0

<210> 1

<211> 957

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> CDS

<222> (1)..(957)

<400> 1

atg ccc gag tat ttg ctt ctg ccc gct ggc cta att tcc ctc tcc ctg	48
Met Pro Glu Tyr Leu Leu Leu Pro Ala Gly Leu Ile Ser Leu Ser Leu	
1 5 10 15	
gcg atc gcc gct gga ctg tat ctc cta act gcc cgg ggc tat cag tca	96
Ala Ile Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr Gln Ser	
20 25 30	
tcg gat tcc gtg gcc aac gcc tac gac caa tgg aca gag gac ggc att	144
Ser Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp Gly Ile	
35 40 45	
ttg gaa tat tac tgg ggc gac cat atc cac ctc ggc cat tat ggc gat	192
Leu Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr Gly Asp	
50 55 60	
ccg cca gtg gcc aag gat ttc atc caa tcg aaa att gat ttt gtc cat	240
Pro Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe Val His	
65 70 75 80	
gcc atg gcc cag tgg ggc gga tta gat aca ctt ccc ccc ggc aca acg	288
Ala Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly Thr Thr	

85	90	95	
gta ttg gat gtg ggt tgc ggc att ggc ggt agc agt cgc att ctc gcc			336
Val Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile Leu Ala			
100	105	110	
aaa gat tat ggt ttt aac gtt acc ggc atc acc att agt ccc caa cag			384
Lys Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro Gln Gln			
115	120	125	
gtg aaa cgg gcg acg gaa tta act cct ccc gat gtg acg gcc aag ttt			432
Val Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala Lys Phe			
130	135	140	
gcg gtg gac gat gct atg gct ttg tct ttt cct gac ggt agt ttc gac			480
Ala Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser Phe Asp			
145	150	155	160
gta gtt tgg tcg gtg gaa gca ggg ccc cac atg cct gac aaa gct gtg			528
Val Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys Ala Val			
165	170	175	
ttt gcc aag gaa tta ctg cgg gtc gtg aaa cca ggg ggc att ctg gtg			576
Phe Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile Leu Val			
180	185	190	
gtg gcg gat tgg aat caa cgg gac gat cgc caa gtg ccc ctc aac ttc			624
Val Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu Asn Phe			
195	200	205	
tgg gaa aaa cca gtg atg cga caa ctg ttg gat caa tgg tcc cac cct			672
Trp Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser His Pro			
210	215	220	
gcc ttt gcc agc att gaa ggt ttt gcg gaa aat ttg gaa gcc acg ggt			720
Ala Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala Thr Gly			
225	230	235	240
ttg gtg gag ggc cag gtg act act gct gat tgg act gta ccg acc ctc			768
Leu Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro Thr Leu			
245	250	255	
ccc gct tgg ttg gat acc att tgg cag ggc att atc cgg ccc cag ggc			816
Pro Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro Gln Gly			
260	265	270	
tgg tta caa tac ggc att cgt ggg ttt atc aaa tcc gtg cgg gaa gta			864
Trp Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg Glu Val			
275	280	285	

ccg act att tta ttg atg cgc ctt gcc ttt ggg gta gga ctt tgt cgc 912
 Pro Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu Cys Arg
 290 295 300

ttc ggt atg ttc aaa gca gtg cga aaa aac gcc act caa gct taa 957
 Phe Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala
 305 310 315

<210> 2

<211> 318

<212> PRT

<213> Synechocystis PCC6803

<400> 2

Met Pro Glu Tyr Leu Leu Leu Pro Ala Gly Leu Ile Ser Leu Ser Leu
 1 5 10 15

Ala Ile Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr Gln Ser
 20 25 30

Ser Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp Gly Ile
 35 40 45

Leu Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr Gly Asp
 50 55 60

Pro Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe Val His
 65 70 75 80

Ala Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly Thr Thr
 85 90 95

Val Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile Leu Ala
 100 105 110

Lys Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro Gln Gln
 115 120 125

Val Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala Lys Phe
 130 135 140

Ala Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser Phe Asp
 145 150 155 160

Val Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys Ala Val
 165 170 175

4

Phe Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile Leu Val
 180 185 190
 Val Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu Asn Phe
 195 200 205
 Trp Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser His Pro
 210 215 220
 Ala Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala Thr Gly
 225 230 235 240
 Leu Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro Thr Leu
 245 250 255
 Pro Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro Gln Gly
 260 265 270
 Trp Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg Glu Val
 275 280 285
 Pro Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu Cys Arg
 290 295 300
 Phe Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala
 305 310 315

<210> 3

<211> 974

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> CDS

<222> (7)..(963)

<400> 3

ggatcc atg ccc gag tat ttg ctt ctg ccc gct ggc cta att tcc ctc 48
 Met Pro Glu Tyr Leu Leu Leu Pro Ala Gly Leu Ile Ser Leu
 1 5 10
 tcc ctg gcg atc gcc gct gga ctg tat ctc cta act gcc cgg ggc tat 96
 Ser Leu Ala Ile Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr
 15 20 25 30
 cag tca tcg gat tcc gtg gcc aac gcc tac gac caa tgg aca gag gac 144
 Gln Ser Ser Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp
 35 40 45

ggc att ttg gaa tat tac tgg ggc gac cat atc cac ctc ggc cat tat	192
Gly Ile Leu Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr	
50 55 60	
ggc gat ccg cca gtg gcc aag gat ttc atc caa tcg aaa att gat ttt	240
Gly Asp Pro Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe	
65 70 75	
gtc cat gcc atg gcc cag tgg ggc gga tta gat aca ctt ccc ccc ggc	288
Val His Ala Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly	
80 85 90	
aca acg gta ttg gat gtg ggt tgc ggc att ggc ggt agc agt cgc att	336
Thr Thr Val Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile	
95 100 105 110	
ctc gcc aaa gat tat ggt ttt aac gtt acc ggc atc acc att agt ccc	384
Leu Ala Lys Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro	
115 120 125	
caa cag gtg aaa cgg gcg acg gaa tta act cct ccc gat gtg acg gcc	432
Gln Gln Val Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala	
130 135 140	
aag ttt gcg gtg gac gat gct atg gct ttg tct ttt cct gac ggt agt	480
Lys Phe Ala Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser	
145 150 155	
ttc gac gta gtt tgg tcg gtg gaa gca ggg ccc cac atg cct gac aaa	528
Phe Asp Val Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys	
160 165 170	
gct gtg ttt gcc aag gaa tta ctg cgg gtc gtg aaa cca ggg ggc att	576
Ala Val Phe Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile	
175 180 185 190	
ctg gtg gtg gcg gat tgg aat caa cgg gac gat cgc caa gtg ccc ctc	624
Leu Val Val Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu	
195 200 205	
aac ttc tgg gaa aaa cca gtg atg cga caa ctg ttg gat caa tgg tcc	672
Asn Phe Trp Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser	
210 215 220	
cac cct gcc ttt gcc agc att gaa ggt ttt gcg gaa aat ttg gaa gcc	720
His Pro Ala Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala	
225 230 235	

6

acg ggt ttg gtg gag ggc cag gtg act act gct gat tgg act gta ccg 768
 Thr Gly Leu Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro
 240 245 250

acc ctc ccc gct tgg ttg gat acc att tgg cag ggc att atc cgg ccc 816
 Thr Leu Pro Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro
 255 260 265 270

cag ggc tgg tta caa tac ggc att cgt ggg ttt atc aaa tcc gtg cgg 864
 Gln Gly Trp Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg
 275 280 285

gaa gta ccg act att tta ttg atg cgc ctt gcc ttt ggg gta gga ctt 912
 Glu Val Pro Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu
 290 295 300

tgt cgc ttc ggt atg ttc aaa gca gtg cga aaa aac gcc act caa gct 960
 Cys Arg Phe Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala
 305 310 315

taa attgcggatc c 974

<210> 4

<211> 318

<212> PRT

<213> Synechocystis PCC6803

<400> 4

Met Pro Glu Tyr Leu Leu Leu Pro Ala Gly Leu Ile Ser Leu Ser Leu
 1 5 10 15

Ala Ile Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr Gln Ser
 20 25 30

Ser Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp Gly Ile
 35 40 45

Leu Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr Gly Asp
 50 55 60

Pro Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe Val His
 65 70 75 80

Ala Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly Thr Thr
 85 90 95

Val Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile Leu Ala

100		105		110
Lys Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro Gln Gln				
115		120		125
Val Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala Lys Phe				
130		135		140
Ala Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser Phe Asp				
145		150		155
Val Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys Ala Val				
	165		170	175
Phe Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile Leu Val				
	180		185	190
Val Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu Asn Phe				
	195		200	205
Trp Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser His Pro				
	210		215	220
Ala Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala Thr Gly				
225		230		235
Leu Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro Thr Leu				
	245		250	255
Pro Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro Gln Gly				
	260		265	270
Trp Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg Glu Val				
	275		280	285
Pro Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu Cys Arg				
	290		295	300
Phe Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala				
305		310		315

<210> 5

<211> 27

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> primer_bind

<222> (1)..(27)

<400> 5

ggatccatgc ccgagtatgt gcttctg

27

<210> 6

<211> 26

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> primer_bind

<222> (1)..(26)

<400> 6

ggatccgcaa tttaagcttg agtggc

26

<210> 7

<211> 930

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> CDS

<222> (10)..(915)

<400> 7

gatatcacc atg gcc gct gga ctg tat ctc cta act gcc cgg ggc tat cag 51

Met Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr Gln

1

5

10

tca tcg gat tcc gtg gcc aac gcc tac gac caa tgg aca gag gac ggc 99

Ser Ser Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp Gly

15

20

25

30

att ttg gaa tat tac tgg ggc gac cat atc cac ctc ggc cat tat ggc 147

Ile Leu Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr Gly

35

40

45

gat ccg cca gtg gcc aag gat ttc atc caa tcg aaa att gat ttt gtc 195

Asp Pro Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe Val

50

55

60

cat gcc atg gcc cag tgg ggc gga tta gat aca ctt ccc ccc ggc aca 243

His Ala Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly Thr

65

70

75

acg gta ttg gat gtg ggt tgc ggc att ggc ggt agc agt cgc att ctc	291
Thr Val Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile Leu	
80 85 90	
gcc aaa gat tat ggt ttt aac gtt acc ggc atc acc att agt ccc caa	339
Ala Lys Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro Gln	
95 100 105 110	
cag gtg aaa cgg gcg acg gaa tta act cct ccc gat gtg acg gcc aag	387
Gln Val Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala Lys	
115 120 125	
ttt gcg gtg gac gat gct atg gct ttg tct ttt cct gac ggt agt ttc	435
Phe Ala Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser Phe	
130 135 140	
gac gta gtt tgg tcg gtg gaa gca ggg ccc cac atg cct gac aaa gct	483
Asp Val Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys Ala	
145 150 155	
gtg ttt gcc aag gaa tta ctg cgg gtc gtg aaa cca ggg ggc att ctg	531
Val Phe Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile Leu	
160 165 170	
gtg gtg gcg gat tgg aat caa cgg gac gat cgc caa gtg ccc ctc aac	579
Val Val Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu Asn	
175 180 185 190	
ttc tgg gaa aaa cca gtg atg cga caa ctg ttg gat caa tgg tcc cac	627
Phe Trp Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser His	
195 200 205	
cct gcc ttt gcc agc att gaa ggt ttt gcg gaa aat ttg gaa gcc acg	675
Pro Ala Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala Thr	
210 215 220	
ggt ttg gtg gag ggc cag gtg act act gct gat tgg act gta ccg acc	723
Gly Leu Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro Thr	
225 230 235	
ctc ccc gct tgg ttg gat acc att tgg cag ggc att atc cgg ccc cag	771
Leu Pro Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro Gln	
240 245 250	
ggc tgg tta caa tac ggc att cgt ggg ttt atc aaa tcc gtg cgg gaa	819
Gly Trp Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg Glu	
255 260 265 270	
gta ccg act att tta ttg atg cgc ctt gcc ttt ggg gta gga ctt tgt	867

10

Val Pro Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu Cys
 275 280 285

cgc ttc ggt atg ttc aaa gca gtg cga aaa aac gcc act caa gct taa 915
 Arg Phe Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala
 290 295 300

attcttaagg tcgac 930

<210> 8

<211> 301

<212> PRT

<213> Synechocystis PCC6803

<400> 8

Met Ala Ala Gly Leu Tyr Leu Leu Thr Ala Arg Gly Tyr Gln Ser Ser
 1 5 10 15

Asp Ser Val Ala Asn Ala Tyr Asp Gln Trp Thr Glu Asp Gly Ile Leu
 20 25 30

Glu Tyr Tyr Trp Gly Asp His Ile His Leu Gly His Tyr Gly Asp Pro
 35 40 45

Pro Val Ala Lys Asp Phe Ile Gln Ser Lys Ile Asp Phe Val His Ala
 50 55 60

Met Ala Gln Trp Gly Gly Leu Asp Thr Leu Pro Pro Gly Thr Thr Val
 65 70 75 80

Leu Asp Val Gly Cys Gly Ile Gly Gly Ser Ser Arg Ile Leu Ala Lys
 85 90 95

Asp Tyr Gly Phe Asn Val Thr Gly Ile Thr Ile Ser Pro Gln Gln Val
 100 105 110

Lys Arg Ala Thr Glu Leu Thr Pro Pro Asp Val Thr Ala Lys Phe Ala
 115 120 125

Val Asp Asp Ala Met Ala Leu Ser Phe Pro Asp Gly Ser Phe Asp Val
 130 135 140

Val Trp Ser Val Glu Ala Gly Pro His Met Pro Asp Lys Ala Val Phe
 145 150 155 160

Ala Lys Glu Leu Leu Arg Val Val Lys Pro Gly Gly Ile Leu Val Val
 165 170 175

11

Ala Asp Trp Asn Gln Arg Asp Asp Arg Gln Val Pro Leu Asn Phe Trp
 180 185 190

Glu Lys Pro Val Met Arg Gln Leu Leu Asp Gln Trp Ser His Pro Ala
 195 200 205

Phe Ala Ser Ile Glu Gly Phe Ala Glu Asn Leu Glu Ala Thr Gly Leu
 210 215 220

Val Glu Gly Gln Val Thr Thr Ala Asp Trp Thr Val Pro Thr Leu Pro
 225 230 235 240

Ala Trp Leu Asp Thr Ile Trp Gln Gly Ile Ile Arg Pro Gln Gly Trp
 245 250 255

Leu Gln Tyr Gly Ile Arg Gly Phe Ile Lys Ser Val Arg Glu Val Pro
 260 265 270

Thr Ile Leu Leu Met Arg Leu Ala Phe Gly Val Gly Leu Cys Arg Phe
 275 280 285

Gly Met Phe Lys Ala Val Arg Lys Asn Ala Thr Gln Ala
 290 295 300

<210> 9

<211> 31

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> primer_bind

<222> (1)..(31)

<400> 9

gatatcacca tggccgctgg actgtatctc c

31

<210> 10

<211> 31

<212> DNA

<213> Synechocystis PCC6803

<220>

<221> primer_bind

<222> (1)..(31)

<400> 10

gtcgacctta agaatttaag cttgagtggc g

31

INTERNATIONAL SEARCH REPORT

International Application No
PCT/EP 00/05862

A. CLASSIFICATION OF SUBJECT MATTER

IPC 7 C12N15/82 C12N15/54 C12N9/10 C12N15/31 G01N33/53
A01H5/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N G01N A01H

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, BIOSIS, PAJ, WPI Data, STRAND

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 31 October 1996 (1996-10-31) KANEKO, T., ET AL. : "sequence analysis of the genome of the unicellular cyanobacterium Syecchocystis sp. PCC6803. II. sequence determination of the entire genome and assignment of the potential protein-coding regions" XP002152668 accession no. D90914</p> <p style="text-align: center;">--- -/--</p>	1

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

* Special categories of cited documents :

- *A* document defining the general state of the art which is not considered to be of particular relevance
- *E* earlier document but published on or after the international filing date
- *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- *O* document referring to an oral disclosure, use, exhibition or other means
- *P* document published prior to the international filing date but later than the priority date claimed

- *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- *G* document member of the same patent family

Date of the actual completion of the international search

22 November 2000

Date of mailing of the international search report

04/12/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Authorized officer

Holtorf, S

INTERNATIONAL SEARCH REPORT

International Application No

PCT/EP 00/05862

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	HOEFGEN R ET AL: "BIOCHEMICAL AND GENETIC ANALYSIS OF DIFFERENT PATATIN ISOFORMS EXPRESSED IN VARIOUS ORGANS OF POTATO SOLANUM-TUBEROSUM" PLANT SCIENCE (LIMERICK), vol. 66, no. 2, 1990, pages 221-230, XP000964790 ISSN: 0168-9452 cited in the application page 223, left-hand column ----	5,6
A	WO 99 04622 A (UNIV NEVADA) 4 February 1999 (1999-02-04) cited in the application ----	
P,X	WO 00 10380 A (UNIV NEVADA) 2 March 2000 (2000-03-02) the whole document ----	1-8
P,X	WO 00 32757 A (RAFALSKI J ANTONI ;DU PONT (US); COUGHLAN SEAN J (US); MIAO GUO HU) 8 June 2000 (2000-06-08) the whole document -----	9,10

INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/EP 00/05862

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9904622 A	04-02-1999	AU 8506198 A EP 1009812 A	16-02-1999 21-06-2000
WO 0010380 A	02-03-2000	AU 5786199 A	14-03-2000
WO 0032757 A	08-06-2000	AU 2037700 A	19-06-2000

INTERNATIONALER RECHERCHENBERICHT

Internationales Aktenzeichen

PCT/EP 00/05862

A. KLASSIFIZIERUNG DES ANMELDUNGSGEGENSTANDES

IPK 7 C12N15/82 C12N15/54 C12N9/10 C12N15/31 G01N33/53
A01H5/00

Nach der Internationalen Patentklassifikation (IPK) oder nach der nationalen Klassifikation und der IPK

B. RECHERCHIERTE GEBIETE

Recherchierter Mindestprüfstoff (Klassifikationssystem und Klassifikationssymbole)

IPK 7 C12N G01N A01H

Recherchierte aber nicht zum Mindestprüfstoff gehörende Veröffentlichungen, soweit diese unter die recherchierten Gebiete fallen

Während der internationalen Recherche konsultierte elektronische Datenbank (Name der Datenbank und evtl. verwendete Suchbegriffe)

EPO-Internal, BIOSIS, PAJ, WPI Data, STRAND

C. ALS WESENTLICH ANGESEHENE UNTERLAGEN

Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	<p>DATABASE EMBL SEQUENCE DATABASE 'Online! 31. Oktober 1996 (1996-10-31) KANEKO, T., ET AL. : "sequence analysis of the genome of the unicellular cyanobacterium Syecchocystis sp. PCC6803. II. sequence determination of the entire genome and assignment of the potential protein-coding regions" XP002152668 accession no. D90914</p> <p style="text-align: center;">--- -/-</p>	1

☒ Weitere Veröffentlichungen sind der Fortsetzung von Feld C zu entnehmen

☒ Siehe Anhang Patentfamilie

* Besondere Kategorien von angegebenen Veröffentlichungen :

A Veröffentlichung, die den allgemeinen Stand der Technik definiert, aber nicht als besonders bedeutsam anzusehen ist

E älteres Dokument, das jedoch erst am oder nach dem internationalen Anmeldedatum veröffentlicht worden ist

L Veröffentlichung, die geeignet ist, einen Prioritätsanspruch zweifelhaft erscheinen zu lassen, oder durch die das Veröffentlichungsdatum einer anderen im Recherchenbericht genannten Veröffentlichung belegt werden soll oder die aus einem anderen besonderen Grund angegeben ist (wie ausgeführt)

O Veröffentlichung, die sich auf eine mündliche Offenbarung, eine Benutzung, eine Ausstellung oder andere Maßnahmen bezieht

P Veröffentlichung, die vor dem internationalen Anmeldedatum, aber nach dem beanspruchten Prioritätsdatum veröffentlicht worden ist

T Spätere Veröffentlichung, die nach dem internationalen Anmeldedatum oder dem Prioritätsdatum veröffentlicht worden ist und mit der Anmeldung nicht kollidiert, sondern nur zum Verständnis des der Erfindung zugrundeliegenden Prinzips oder der ihr zugrundeliegenden Theorie angegeben ist

X Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann allein aufgrund dieser Veröffentlichung nicht als neu oder auf erfinderischer Tätigkeit beruhend betrachtet werden

Y Veröffentlichung von besonderer Bedeutung; die beanspruchte Erfindung kann nicht als auf erfinderischer Tätigkeit beruhend betrachtet werden, wenn die Veröffentlichung mit einer oder mehreren anderen Veröffentlichungen dieser Kategorie in Verbindung gebracht wird und diese Verbindung für einen Fachmann naheliegend ist

Z Veröffentlichung, die Mitglied derselben Patentfamilie ist

Datum des Abschlusses der internationalen Recherche

22. November 2000

Absendedatum des internationalen Recherchenberichts

04/12/2000

Name und Postanschrift der Internationalen Recherchenbehörde
Europäisches Patentamt, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016

Bevollmächtigter Bediensteter

Holtorf, S

C.(Fortsetzung) ALS WESENTLICH ANGESEHENE UNTERLAGEN		
Kategorie*	Bezeichnung der Veröffentlichung, soweit erforderlich unter Angabe der in Betracht kommenden Teile	Betr. Anspruch Nr.
X	HOEFGEN R ET AL: "BIOCHEMICAL AND GENETIC ANALYSIS OF DIFFERENT PATATIN ISOFORMS EXPRESSED IN VARIOUS ORGANS OF POTATO SOLANUM-TUBEROSUM" PLANT SCIENCE (LIMERICK), Bd. 66, Nr. 2, 1990, Seiten 221-230, XP000964790 ISSN: 0168-9452 in der Anmeldung erwähnt Seite 223, linke Spalte ---	5,6
A	WO 99 04622 A (UNIV NEVADA) 4. Februar 1999 (1999-02-04) in der Anmeldung erwähnt ---	
P,X	WO 00 10380 A (UNIV NEVADA) 2. März 2000 (2000-03-02) das ganze Dokument ---	1-8
P,X	WO 00 32757 A (RAFALSKI J ANTONI ;DU PONT (US); COUGHLAN SEAN J (US); MIAO GUO HU) 8. Juni 2000 (2000-06-08) das ganze Dokument -----	9,10

INTERNATIONALER RECHERCHENBERICHT

Angaben zu Veröffentlichungen, die zur selben Patentfamilie gehören

Intern. altes Aktenzeichen

PCT/EP 00/05862

Im Recherchenbericht angeführtes Patentdokument		Datum der Veröffentlichung	Mitglied(er) der Patentfamilie		Datum der Veröffentlichung
WO 9904622	A	04-02-1999	AU	8506198 A	16-02-1999
			EP	1009812 A	21-06-2000
WO 0010380	A	02-03-2000	AU	5786199 A	14-03-2000
WO 0032757	A	08-06-2000	AU	2037700 A	19-06-2000

SEQUENCE OF THE AMPLIFIED 209 bp PCR-PRODUCT OF THE β -GLOBIN GENEFORWARD PRIMER: β 2

CATTTGCTTC TGACACAAC GTGTCTACTA GCAACCTCAA ACAGACACCA

12mer PRIMER

TGGTGCACCT GACTGCTGTG GAGAAGTCTG CCGTTACTGC CCTGTGGGGC

AAGGTGAACG TGGATGAAGT TGGTGGTGAG GCCCTGGGCA GGTGGGTATC

AAGGTTACAA GACAGGTTTA AGGAGACCAA TAGAACTGG GCATGTGGAG

ACAGAGAAG

REVERSE PRIMER β 11*FIG. 51*